(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 February 2003 (27.02.2003)

PCT

(10) International Publication Number
WO 03/016875 A2

(51) International Patent Classification[7]: G01N 15/14

(21) International Application Number: PCT/US02/26334

(22) International Filing Date: 19 August 2002 (19.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/932,413 17 August 2001 (17.08.2001) US
10/076,363 15 February 2002 (15.02.2002) US

(71) Applicant *(for all designated States except US)*: **UNION BIOMETRICA, INC.** [US/US]; 35 Medford Street, Suite 213, Somerville, MA 02143 (US).

(72) Inventors; and

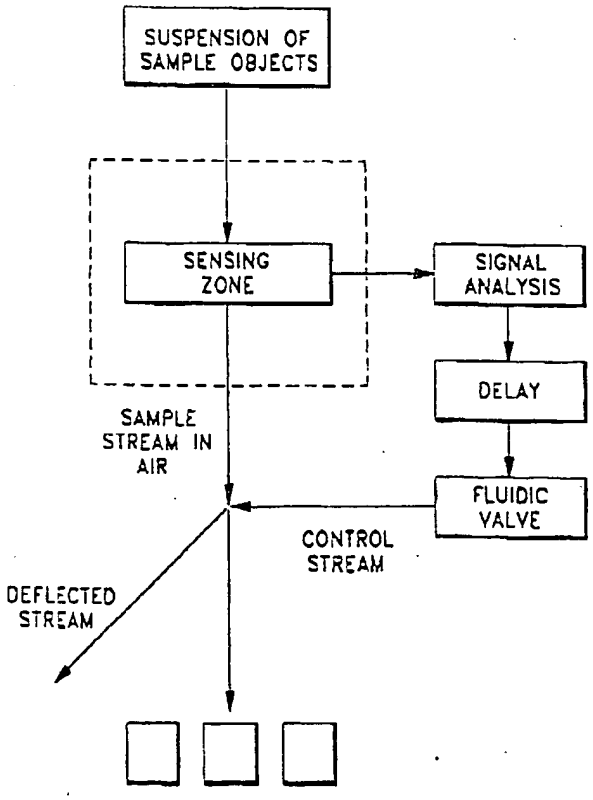
(75) Inventors/Applicants *(for US only)*: **HANSEN, W., Peter** [US/US]; 121 Top of Dean Hill Road, Canaan, NY 12029 (US). **FERRANTE, Anthony, A.** [US/US]; 15 Westlund Road, Belmont, MA 02178 (US). **PERRAULT, Donald, F.** [US/US]; 25 Bennett Street, Brighton, MA 02135 (US). **GERSHMAN, Russell, J.** [US/US]; 19 Ward Street, Somerville, MA 02143 (US). **KRAULEDAT, Petra, B.** [US/US]; 121 Top of Dean Hill Road, New York, NY 12029 (US). **PULAK, Rock** [US/US]; 55 Warwick Road, Belmont, MA 02178 (US).

(74) Agent: **HUNTER BAKER, C.**; Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

[Continued on next page]

(54) Title: SYSTEM FOR AXIAL PATTERN ANALYSIS OF MULTICELLULAR ORGANISMS



(57) Abstract: This invention relates to a method of using elongate multicellular organisms in conjunction with a specialized flow cytometer for drug discovery and compound screening. A stable, optically detectable linear marker pattern on each organism is used to construct a longitudinal map of each organism as it passes through the analysis region of the flow cytometer. This pattern is used to select and sort multicellular organisms based on particular phenotypic characteristics. The invention further combines the selection and sorting aspects with an identification aspect so that the genotype of each multicellular organism can be identified.

WO 03/016875 A2

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SYSTEM FOR AXIAL PATTERN ANALYSIS OF MULTICELLULAR ORGANISMS

Priority Information

This application is a continuation-in-part of U.S. Application No. 09/932,413, filed August 17, 2001, which claims priority to U.S. Provisional Application No. 60/226,701, filed August 18, 2000. This application is also a continuation in part of U.S. Application No. 09/465,215, filed December 15, 1999, which claims priority to U.S. Provisional Application No. 60/112,280, filed December 15, 1998. This application is also a continuation-in-part of U.S. Application No. 10/076,363 filed February 15, 2002. Each of the above-referenced applications is incorporated herein by reference.

Field of the Invention

This invention concerns the field of automated analysis of complex, multicellular model organisms that are particularly useful in the field of drug discovery, of toxicology, and development.

Background of the Invention

Rapid, high-throughput compound screening assays have revolutionized the field of drug discovery. Automated drug discovery assays measure changes in a variety of cells grown *in vitro*. For example, microarray assays simultaneously assess the effect of thousands of compounds on a particular biochemical pathway *in vitro*. However, such automated drug screening assays are not readily available for assessing the *in vivo* effects of multiple compounds on complex multicellular organisms. In particular, assessing the effect of one or more compounds on the development or physiology of a multicellular organism remains a tedious manual task requiring hundreds of hours of labor by highly skilled technicians.

Intact multicellular organisms, such as the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, or the zebrafish *Danio rerio* are frequently used as model systems to help understand the function of human genes that have been implicated in disease. Human gene homologs are often identified in these model

organisms and provide valuable tools for studying the biological function of the genes *in vivo*. Such mutations frequently result in an easily observable phenotypic change in the model organism. Furthermore, it has been shown that certain pharmacological compounds collaterally produce optically detectable changes in these mutant organisms. These changes can be used to identify specific compounds that interact with a particular gene product *in vivo*. The addition of such functional genomic techniques to the repertoire of molecular biology and biochemistry methods can greatly accelerate the drug discovery process.

Mutants of intact organisms are used as a new class of methods for *in vivo* drug screening of libraries of potential pharmacological compound produced through the use of combinatorial chemical methods. With these organisms, one can identify targets for drug intervention without the need to completely understand the complex biochemical pathways that relate the genome to the phenotype. In addition, investigators can annotate drug libraries for toxicity, non-specific activity, or cell membrane permeability by observing their behavior in intact organisms. In this way, toxic or ineffective libraries and/or library members can be discarded at an early stage of testing without wasting valuable resources. This allows rapid and economical screenings of the compound libraries for new and useful therapeutic compounds, while limiting politically controversial testing on mammals.

While model organisms such as *C. elegans, D. melanogaster,* and *D. rerio* have been proven useful in the study of human disease, they have not yet been successfully used in the field of high speed, high throughput drug discovery. This presents a significant hindrance to investigators that need to search through thousands of multicellular organisms for the phenotype of a new mutation or for a response to a panel of sample drugs. For example, with today's molecular biology techniques, a large laboratory can produce deletion mutations in a multicellular test organism at a rate of 20 to 30 per month. In order to evaluate the effect of each member of a chemical compound library (that frequently contains 100,000 discrete compounds) on a single deletion mutant, one must manually manipulate and deposit a precise number of organisms of the mutant strain at the same developmental stage into various containers, such as wells of a microtiter plate array. Wild type or deviants from the desired mutant strain, or organisms at a different development stage must be

eliminated. The use of such slow, manual methods for the selection and deposition of organisms of the proper type greatly delays the entire process of drug discovery. Moreover, manual methods rely on pipettes that dispense accurate volumes of fluid but not accurate numbers of organisms. In many studies, where reproduction rate is altered by the mutation, it is necessary to begin with an exact and known number of multicellular organisms in each well. This is, at best, a daunting requirement.

The effect of a therapeutic compound or toxic environment on the mutant strain or expression system can be determined by identifying changes in the spatial pattern of fluorescence or staining. Fluorescent protein genes are typically used as reporters for gene expression in a wide variety of organisms (Tsien , R. *Nature Biotechnology* (1999) 17: 956-57). For example, green fluorescent protein (GFP) is used as a reporter gene to indicate that an inserted gene has been expressed. The expression of the fluorescent protein usually occurs in a specific spatial pattern within a multicellular organism. Discrimination of one pattern from another is currently carried out manually using a fluorescent microscope. Like the selection and deposition step, this is an extremely tedious task requiring a significant number of workers that are trained at high academic levels.

Prior art methods of selecting multicellular organisms have relied on instruments that performed a "slit-scan" of whole organisms as they passed through the analysis zone of a laser. Methods of detecting fine detail in slit-scanning have relied on apparatuses that utilize diffraction limited optics to create narrow line focus and image plane masks to act as optical spatial filters. This narrow line of focus is sufficient for analyzing single cells, but is insufficient for detecting and spatially locating a particular feature against the more complex background profile of light scatter and autofluorescence presented by a multicellular organism. For example, the diameter of a mature *C. elegans* is approximately 70 micrometers. This means that the background autofluorescence from a nematode is approximately ten times that from a white blood cell (about seven micrometers in diameter), while a fluorescence reporter signal from a single *C. elegans* cell is no greater than that from a single blood cell. In the case of *D. melanogaster* (fruit fly) larvae, the situation is even worse because the diameter of an advanced stage larva is of the order of one millimeter, which means that autofluorescence is much more than a hundred times greater than in

single blood cells. This means that experimentally created, fluorescent features along the axis of a multicellular organism may produce a much weaker optical signal than the autofluorescence background. One can imagine that an axial profile of auto fluorescence with very high peaks and valleys would effectively mask an experimentally created fluorescence feature.

Flow instruments have been used to count the number of nematodes in a fluid volume. Byerly et al (Byerly,et al., *Rev. Sci. Instrum.* (1975) May 46(5): 517-22), described such a device where a flow cytometer employed sheath flow to orient nematodes along the direction of flow so that their size could be measured and organism-by-organism counts could be made by an electrical impedance method. The device was similar to a commercial Coulter counter. The use of the impedance sensor, which can only estimate overall size, and cannot spatially resolve localized features along the major axis of the organism limits the Byerly et al. instrument. In addition to this limitation, the Byerly et al. instrument could not select and deposit (sort) specific organisms.

Co-pending United States Patent Application Serial No. 09/378,634, filed 20-Auguat-1999, which is incorporated herein by reference, describes an instrumentation system for the rapid analysis and sorting of multicellular organisms using optical characteristics such as light scatter and fluorescence to classify each organism in a flowing stream. A single value of fluorescence intensity at a given emission wavelength is detected and assigned to each organism. However, this instrument reports only the intensity, not the position of fluorescence along the major (long) axis of the organisms.

An optical flow instrument for analyzing elongate organisms such as plankton with widths of 500 □m and lengths over 1000 □m has also been described with sheath flow to achieve orientation of the plankton. (J.C. Peters, G.B. Dubelaar, J. Ringelberg, and J.W. Visser, "Optical Plankton Analyser: a Flow Cytometer for Plankton Analysis, I: Design Considerations" Cytometry 1989 Sept 10 (5): 522-528; and G.B. Dubelaar, A.C. Groenwegen, W. Stokdijk, G.J. van den Engh, and J.W. Visser, "Optical Plankton Analyser: a Flow Cytometer for Plankton Analysis, II: Specifications", Cytometry 1989 Sept 10 (5): 529-539). The size range of the plankton used in these optical flow cytometers is similar to that encountered with *C.*

elegans nematodes, fruit fly larvae, and zebrafish embryos; however, there is no provision that avoids the ambiguous light scatter signals that are eliminated by the present invention.

There exists the need for a high-speed system for automatically identifying and physically selecting multicellular organisms with certain spatially distinct, optically detectable, phenotypic characteristics from mixed populations. Such a system must have the ability to locate and measure the intensity and position of experimentally created optical features in the presence of overwhelming autofluorescence.

Summary of the Invention

The present invention provides a system for physically sorting complex multicellular organisms from mixed populations on the basis of spatially distinct, optically detectable, phenotypic characteristics and identifying the genotype of the multicellular organisms once sorted. Two key components of this system include 1) the use of fluorescence markers to provide positional and orientational information about model multicellular organisms and 2) the use of unique identifiers that encode the genotype of the particular multicellular organisms.

According to certain aspects of the invention, strains of model multicellular organisms are generated that have a first marker pattern feature that is spatially consistent and is used to orient the organisms along their major axes. In preferred embodiments, the strains of model organisms also include a second inducible or modifiable feature whose position may be determined in relation to the first marker pattern features. For example, strains of organisms may be generated that include fluorescent marker proteins in one or more particular cells or physiological locations (e.g., the head and the tail) in the organisms. The strains of model organisms may also include a second feature that is inducible by a particular signal, e.g., a genetic mutation. The multicellular organisms of the invention are particularly useful for high-throughput drug screening assays wherein the organisms are exposed to compounds from a combinatorial chemical compound library and assessed for the presence of the particular second feature.

According to the phenotypic analysis aspect of the invention, novel flow cytometry and sorting instrumentation is provided for generating spatial profiles of the model multicellular organisms, which preferably contain the spatially distinct, optically detectable, phenotypic characteristics described herein, by measuring fluorescent and light scattering properties. The model organisms are analyzed and sorted based on spatial profile optical measurements. The technical details of the optical, electronic, and fluidic components of the instrumentation and algorithms of the invention, are described herein.

In related embodiments, the present invention provides two distinct methods of optically detecting an organism in the detector beam of the inventive flow cytometer device. The difference between the methods hinges on the type of second signal used to gate the fluorescence signals on the organisms so that they may be detected over background autofluorescence and noise signals. In one preferred embodiment, light scattered in the forward direction by the organism is used to generate the gate signal for fluorescence detection. In another preferred embodiment, light attenuated by the organism in the forward direction is used to generate a gate signal for fluorescence detection.

A second component of the invention is the ability to identify the genotypes of the multicellular organisms, which are selected based on their phenotypes. This involves detecting a unique identifier that represents the genotype of a particular multicellular organism. In certain preferred embodiments, the unique identifier is a nucleic acid sequence that can be amplified and detected. In other preferred embodiments, the unique identifier is a detectable signal, such as a detectable colorimetric signal or fluorimetric signal. The unique identifier is preferably detectable by the automated flow cytometry and sorting instrument so that the phenotypic sorting and genotypic identifying steps can occur concurrently. Alternatively, detection of the unique identifier occurs after the phenotypic selection.

Experimental data is provided that demonstrates use of the inventive instrument to screen and sort the model organisms having spatially distinct, optically detectable, phenotypic characteristics. This data illustrates the elegant nature of the invention, which combines the use of multicellular organisms having spatially distinct, optically detectable, phenotypic characteristics with an instrument that can

analyze this information and use it to sort the multicellular organisms base on specific spatially distinct features.

Brief Description of the Drawing

The invention is described with reference to the several figures of the drawing, as follows.

Figure 1 shows a diagrammatic representation of optics, flow cell, command electronics, and fluid switch.

Figure 2 shows a diagrammatic representation of the optical beams of the instrument of Figure 1

Figure 3, panels A and B, show diagrams relating fluorescence signals (gated by one of the methods of the invention) related to hermaphroditic (panel 3A) and male (panel 3B) *C. elegans* as measured by the instrument of the present invention.

Figure 4 shows actual oscilloscope traces from a NACLS forward light scatter detector (lower trace) placed at a 45 degree forward light scatter angle and fraction of a degree below the optical axis and a fluorescence detector (upper trace) at right angles to the optic axis.

Figure 5 shows actual oscilloscope traces from a NACLS forward light scatter detector (lower trace) placed at 45 degrees from the optical axis and a fluorescence detector (upper trace) at right angles to the optic axis.

Figure 6 shows actual oscilloscope traces from an extinction detector (lower trace) placed on the optical axis and a fluorescence detector at right angles to the optical axis (upper trace).

Figure 7 shows actual oscilloscope traces from a WACLS forward light scatter detector (lower trace) and a fluorescence detector at right angles to the optical axis (upper trace); the *C. elegans* samples scanned showed several discreet points of fluorescence.

Figure 8 shows actual oscilloscope traces from a WACLS forward light scatter detector (lower trace) and a fluorescence detector at right angles to the optical axis (upper trace); the *C. elegans* specimens scanned showed a small additional fluorescence at one end.

Figure 9 shows a series of photomicrographs of transgenic *C. elegans* expressing ZsYellow under the control of the *egl-17* promoter with panel A showing a white light image with the corresponding fluorescence image shown in panel C; panel B similarly corresponds to the fluorescence image of panel D.

Figure 10 shows oscilloscope tracings of optical detector signals resulting from flow cytometric analysis of *C. elegans* with panel A and panel B showing transgenic *egl-17* expressing organisms corresponding to the organisms of Figure 3 and panel C showing wild type *C. elegans* as a control.

Figure 11 shows PY1089 transgenic organisms expressing GFP (and showing autofluorescence) with panel A showing a light micrograph and panel B showing the corresponding fluorescence micrograph, while panel C and panel D show optical detector signals from these organisms undergoing flow analysis.

Figure 12 shows optical detector signals that result from flow cytometric analysis of transgenic organisms resulting from mating the *egl-17::ZsYellow* construct into PY1089.

Figure 13 illustrates the instrument of the invention that is a flow cytometer equipped to sort multicellular organisms based on spatially distinct, optically detectable, phenotypic characteristics.

Figure 14 illustrates the arrangement of the standard emissions filters in the inventive sorting device.

Figure 15 illustrates the basic concept of the inventive sorting device where organisms that do not display a desired spatially distinct, optically detectable, phenotypic characteristic are separated from the organisms that do have a desired spatially distinct, optically detectable, phenotypic characteristic.

Figure 16 is a graph showing a trace of a worm expressing the ZsYellow protein using the *egl*-17 *C. elegans* promoter. Expression is detected in one head neuron (left side of trace) and in the vulva (middle peak).

Figure 17 is a graph showing a trace of a worm expressing the ZsYellow protein measured by an oscilloscope. The axial light loss signal is indicated by the lower trace and the yellow fluorescent signal is indicated by the upper trace.

Figure 18 is a photomicrograph of a transgenic *C. elegans* expressing ZsYellow under control of the *egl*-17 promoter. The upper panel shows a white light

image and the lower panel is the corresponding fluorescence image. The fluorescent image has two bright spots corresponding to the vulva (middle) and the head (right) of the organism.

Figure 19 illustrates upstream and downstream universal primers flanking a nucleic acid identifier sequence.

Figure 20 illustrates the results of hybridizing amplified nucleic acid identifiers to a DNA chip containing complementary nucleic acid sequences.

Detailed Description of the Invention

Drug screening assays generally focus on a single step of an often complex biochemical pathway. A compound that affects the biochemical pathway is generally identified as a "hit." Assays can be designed in complex, living, multicellular organisms such that a compound that affects any component of a biochemical pathway is identified as a "hit." Use of a complex, multicellular organism can provide data relating to the toxicity and the impact of a compound on specific biochemical events. The present invention provides, for the first time, an automated system for identifying and physically selecting multicellular organisms from mixed populations on the basis of spatially distinct, optically detectable, phenotypic characteristics. The present invention further combines the automated identification and selection system with a system for identifying the genotype of the selected multicellular organisms.

By including an automated instrument, namely a flow analyzer and sorter, and a population of organisms with marker pattern features, the system of the present invention allows one to localize and report not only the intensity of fluorescence on an organism, but also the position of fluorescence along the major (long) axis of an organism. The inventive automated system, generating new spatial information can be used to separate (sort) mutant organisms with a particular phenotypic trait from a mixed population of organisms. In certain preferred embodiments, the particular phenotypic trait can result from contact of the organism with a compound. In other preferred embodiments, the particular phenotypic trait results from genetically crossing two genotypically different multicellular organisms. Alternatively, the automated system of the present invention can be used to separate multicellular

organisms that are at a particular stage of development. Examples of applicable multicellular organisms are all stages developmental of *Caenorhabditis elegans*; *Drosophila melanogaster* (fruit fly) eggs, larva, and embryos; or *Danio rerio* (zebrafish) eggs, larva, and embryos, which are useful as model organisms for human disease and functional genomics studies. According to certain preferred embodiments, organisms, such as nematodes, having markers at every stage of development can be generated for use in the present invention. Specific stages of certain ascidians and amphibians can also be analyzed using the inventive systems and methods. Other organisms that can be used with the invention include certain life cycle stages of certain Coelenterates, Platyhelminthes, Rotifers, Mollusks, other Nematodes, Echinoderms, Arthropods, other Hemichordates, and Chordates.

According to a particularly preferred embodiment of present invention, any stage of any non-human multicellular animal can be profiled according to the present invention. For examples, embryos of any non-human multicellular animal that fall within a certain size range can be profiled by the present invention. In one preferred embodiment, any multicellular organism undergoing a developmental stage in which the multicellular organism falls within the size range of 25 micrometers to 1,250 micrometers can be profiled according to the present invention (see, e.g., Barnes, Robert D. Invertebrate Zoology, W.B. Saunders Co. Philadelphia, London, Toronto, incorporated herein by reference).

In related embodiments, the system can be used to profile any elongate object within the size range of 25 micrometers to 1250 micrometers in diameter. For example, beads falling within 25 micrometers to 1,250 micrometers, i.e., used in bead chemistry applications such as combinatorial chemistry applications, can be selected and sorted according to the present invention using any of a variety of well known conventional combinatorial on bead methods. According to the present invention, the beads include at least one second feature that is modifiable or inducible by a test treatment to produce a particular characteristic that can form the basis for sorting the elongate object. The second feature can be for example, any of the detectable signals or signal producing systems described herein or known in the art. For example, beads containing a potential ligand for a protein of interest can be sorted based on the detection of a fluorescent signal attached to a bound protein of interest. According to

certain preferred embodiments at least one detectable signal can be used to identify an agent attached to an object by attaching the detectable signal to the object. In addition, plant seeds, e.g., those of Arabidopsis thaliana, can be sorted using this system.

The inventive system, which combines strains of multicellular organisms characterized by a stable spatial pattern of fluorescence, staining, or other optically detectable characteristics, with an instrument that can accurately sort multicellular organisms based on the position of an experimental feature relative to other invariant features by axial scanning, greatly benefits the areas of developmental biology and drug discovery. As pointed out above, multicellular organisms are frequently used as models for human disease and provide excellent tools for testing the effect of a test compound on a particular biochemical pathway *in vivo*.

For example, the effect of a therapeutic compound or toxic environment *in vivo* can be determined by monitoring changes in the spatial patterns of selected features on the multicellular organisms. For example, in analyzing the effects of a drug screen on a particular feature, it would be desirable to know if the feature was present or absent, or up-regulated or down-regulated. As but another example, locating spatial features on a multicellular organism is important for understanding development and differentiation of structures during the organism's life cycle. Using the system of the present invention, one could also tag a cell for abnormal cell migration. One could examine a particular *in vivo* function, e.g., axon growth (e.g., by measuring the length of an axon by labeling the axon cell with a fluorescent marker) and determine whether reversing the mutation causes the nerve to grow to a normal length.

The invention further provides systems for identifying strains of the phenotypically selected multicellular organisms. In one preferred embodiment, each multicellular organism has a unique identifier that specifies the genotype of the strain. In certain preferred embodiments, the unique identifier is a unique nucleic acid sequence that can be amplified and detected. In other preferred embodiments, the unique identifier is a detectable signal that is preferably detected directly by the automated instrument. In certain preferred embodiments, phenotypic and genotypic

identification happen simultaneously. Alternatively, the unique identifier may be detected after phenotypic selection has been completed.

Automating the detection of spatial patterns of signals further improves the objectivity and speed of analysis. The present invention provides the first system that is suitable for the high-speed analysis and placement of axial features within transparent, or partially transparent, multicellular organisms. The present invention also provides the first system that is capable of performing a phenotypic selection on a large population of multicellular organisms all at once, avoiding the tedious manual labor typically required for screening such large populations. Such a system has wide applicability in academic research and pharmaceutical applications. The components of this unique system, including the multicellular organisms, the instrument, and the identification system are discussed in detail below.

Fluorescent Markers for Spatial Analysis

The system of the present invention includes a population of multicellular organisms having a plurality of spatially distinct, optically detectable, phenotypic characteristics and an instrument for detecting such characteristics. According to the present invention, a spatially distinct, optically detectable, phenotypic characteristic includes a marker pattern of spatially consistent first features that can be used to orient the organism along its major axis. The spatially distinct, optically detectable, phenotypic characteristics may also include a second feature that is modifiable or inducible when the population is subjected to a test treatment. Test treatments include exposure to a chemical compound (e.g., a chemical compound from a complex chemical compound library), a harsh environment (e.g., starvation, increased temperature, decreased temperature, overcrowding, changes in light etc.), a mutagenesis procedure (e.g., exposure to a known chemical mutagen), a genetic crossing procedure (e.g., where two strains of organisms are crossed to determine how particular phenotypic traits or biochemical pathways interact), etc.

Those skilled in the art will recognize that any spatially distinct, optically detectable, phenotypic characteristics may be used to orient an organism along its major axis. In certain preferred embodiments, the organism is aligned using only the first marker pattern of spatially consistent features and does not utilize the second

feature induced by the test treatment, e.g., if one were to merely separate organisms with a particular set of first features.

The present invention further provides methods for sorting multicellular organisms based on a particular phenotype. A population is provided wherein each member of the population displays a marker pattern of spatially distinct first features, which can be used to orient the organism in the detection instrument, and a second feature, whose presence and location is used to identify how the organism is to be sorted. Members of the population of organisms that display the second feature at the proper intensity or location with respect to the features of the marker pattern is separated from the members of the mixed population that do not have the second feature. The arrangement of the second feature with respect to the first features of the marker pattern is a key to the ability to sort the organisms based on specific phenotypic traits.

For example, multicellular organisms can be sorted by first providing a population of test organisms having any combination of first marker pattern features and second features; measuring the arrangement of the first marker pattern features and the second features; and depositing members of the population based on the arrangement of the second feature with respect to the first marker pattern features.

In certain preferred embodiments, a plurality of spatially distinct, optically detectable, phenotypic characteristics also includes multiple second features that are modifiable or inducible. First and second features can range in size from 10 micrometers to the full length of the organism. The arrangement of the multiple second features with respect to the first features of the marker pattern can be used to further characterize and sort certain members of the population based on multiple phenotypic characteristics. For example, an organism may have a plurality of second features that are mutations resulting from a mutagenesis procedure, which are detectable by the instrument and can be used to separate the organisms from the mixed population.

In certain preferred embodiments, members of the population are passed through the instrument and sorted based on the arrangement of the first marker pattern features with respect to the one or more second features. Examples of features that can be classified as spatially distinct, optical characteristics include: the localized

expression of DNA encoded fluorescent protein molecules, localized variations of the index of refraction or granularity, and localized variations in specific binding sites (receptors) for optically labeled antibodies, lectins, or other specific ligands. More specific features include the location of specific cells at particular developmental stages or the presence or absence of particular gene products at particular developmental stages, which can each be marked by the presence or absence of a fluorescent protein molecule.

Fluorescent protein genes have been used as reporters for gene expression in a wide variety of organisms (Tsien, R., *Nature Biotechnology* (1999)17: 956-57). The present invention permanently incorporates fluorescent proteins into large populations of multicellular organisms to create spatially marked strains that can be used in combination with a high-speed flow cytometer to detect and map the spatial location of other, experimentally generated second features, e.g., gene expression, with a high degree of precision. The marker patterns serve as guides to focus and synchronize the signal processing and computational electronics on specific spatial regions of the experimental organism where expression is expected, thus improving processing speed and accuracy in
the marked strain, thus providing clues about the developmental aspects of the expression event.

Identification Systems and Methods

Generally, the present invention provides a system that combines an automated expression profiling system, which selects multicellular organisms having a particular phenotype, with a system that identifies the genotype of each selected multicellular organism.. In preferred embodiments, the present invention provides systems and methods for identifying genetically engineered multicellular organisms having a particular phenotypic characteristic from a population of genetically engineered multicellular organisms.

The inventive system includes a population of genetically engineered multicellular organisms comprising at least one, preferably more than one (i.e., a plurality) of spatially distinct, optically detectable, phenotypic characteristics; an instrument for detecting the location of a spatially distinct, optically detectable,

phenotypic characteristic on the multicellular organism (and for orienting the multicellular organism along its longitudinal axis) and sorting multicellular organisms having a particular phenotypic characteristic; and an identification system for identifying the genotype of each sorted multicellular organism having the particular phenotypic characteristic.

The inventive methods include the steps of 1) providing a population of genetically engineered multicellular organisms; 2) analyzing the arrangement of spatially distinct, optically detectable, phenotypic characteristics on the multicellular organism, wherein the instrument is capable of orienting the organism along its longitudinal axis and sorting multicellular organisms having a particular phenotypic characteristic; and 3) identifying the genotype of the phenotypically sorted multicellular organisms. According to the invention, any system that identifies the genotype of a multicellular organism in the collection of phenotypically selected multicellular organisms may be used.

By combining an automated phenotype selection system with a genotype identification system, the genotypic identity of phenotypically sorted multicellular organisms can be rapidly determined. This is particularly useful for screening mixed populations of multicellular organisms. Instead of ending up with a population of phenotypically sorted organisms whose identity is unknown, the present invention provides a complete selection and identification system that provides both the phenotype and the genotype of the sorted strains.

In preferred embodiments, the present invention provides a system for identifying the genotypes of phenotypically selected multicellular organisms having a particular expression profile. Preferably, the starting population of organisms is a mixed population of strains having different genotypes. As used herein, "genotype" refers to all or part of the genetic constitution of an individual multicellular organism. According to the present invention, the genetic constitution includes any genetic information, whether it is genomic or non-genomic (e.g., plasmid). For example, according to the invention, a mixed population of organisms includes at least 2 strains, at least 10 strains, at least 100 strains, at least 1000 strains, between 1000-10,000 strains, or between 10,000 and 20, 000 strains, wherein each strain has a different genotype.

According to the present invention, a mixed population of strains shares a uniform expression profile, or "baseline" expression profile. The baseline expression profile can be a uniform level, timing, or location of expression of a particular second feature, which is potentially modifiable or inducible by a test treatment. Multiple strains that share a particular baseline expression profile may be identified before exposing the mixed population to one or more test treatments. Each mixed population having a particular baseline expression profile will be exposed to at least one, preferably more than one, test treatment. In related embodiments, multiple mixed populations (each containing multiple strains) that each fit a different baseline expression profile may be used to screen each test treatment.

According to the present invention, a mixed population, of genetically engineered multicellular organisms is subjected to a test treatment, wherein each member of the population displays at least one second feature having a baseline expression profile that may or may not be sensitive to the test treatment. The flow cytometer is equipped to detect changes is the baseline expression profile of the second feature and to sort multicellular organisms having a particular change in the baseline expression profile of the second feature in response to the test treatment. The genotypes of the sorted members of the population are then identified as described herein.

In certain preferred embodiments, at least one, preferably more than one, genomic promoter originating from the multicellular organism to be used in the selection is isolated and placed upstream of a reporter gene encoding a fluorescent protein so that the promoter drives expression of the fluorescent protein. If more than one promoter is used to generate different strains of multicellular organisms, the same reporter gene is preferably used in each strain. In certain preferred embodiments at least two different fluorescent proteins, driven by two different promoters, are expressed simultaneously in a given strain. This two-promoter system facilitates analyzing the effect of a particular test treatment on an intracellular biochemical pathway involving different enzymes that interact with one another. The two-promoter system may also be used to identify new proteins, as well as the genes encoding the proteins, that interact in the pathway. This system is particularly advantageous if the expression profile of multiple genes in the pathway is known.

As described herein, the phenotypic characteristic that is used to select organisms can generally be any change in the baseline expression profile of the second feature(s). For example, changes in the baseline expression profile include an increase or decrease in the level of expression of the second feature; an increase, decrease, or disappearance of expression of the second feature at a specific time during development; and an increase, decrease, or disappearance of expression of the second feature at a specific location during development of the multicellular organism. Alterations in phenotype further include changes in the position, intensity, width, and developmental stage of the baseline expression profile of the second feature in response to a test treatment. It is possible that particular test treatments may mimic or induce certain disease states in the organism, providing useful tools for research and the potential to identify therapeutic compounds. It is possible that particular mutant stain backgrounds may mimic or induce certain disease states in the organism and that these models of disease states be monitored using the optical phenotypes and instrumentation, providing useful tools for research and potential to identify therapeutic compounds.

In preferred embodiments, the identification system of the invention utilizes a unique identifier for each different strain of genetically engineered multicellular organisms screened using the phenotypic selection process. The unique identifier represents the genotype of the organism. According to the present invention, detecting the unique identifier identifies the genotype of the sorted multicellular organism.

Unique identifiers of the present invention include any molecule that has the ability to identify the genotype of the selected strain. Unique identifiers of the present invention include molecules that emit detectable signals (referred to herein as "detectable signals"), e.g., colorimetric signals (e.g., dyes, fluorescent signals or tags, luminescent markers, quantum dots, etc.), radioactive signals, chemical signals (e.g., mass tags, IR tags, UV tags, potentiometric tags, photoactive groups, chelator groups, etc.), polypeptide signals (e.g., enzymatic proteins, e.g., reporter genes, antigens, etc.), and the like.

In certain preferred embodiments, the instrument that is capable of detecting and sorting the multicellular organisms based on their phenotypes is further equipped

to detect the detectable signal generated by the unique identifier. For example, the instrument may include an apparatus that can detect the detectable signal emitted by the unique identifier, thereby identifying the genotype of the sorted organism. According to the present invention, detection of the unique identifier can occur simultaneously with or subsequent to detection of the phenotypic characteristics. In other preferred embodiments, the unique identifier is detected after the multicellular organisms are sorted. In yet other preferred embodiments, the unique identifier is detected before the multicellular organisms are sorted. The analysis of the population of multicellular organism can occur in batch with subsequent separation to multiwell plates for genotype determination. Also, analysis of individual populations can occur from multiwell plates with subsequent separation to multiwell plates for subsequent genotype determination.

One particularly preferred unique identifier is a unique nucleic acid identifier. According to this aspect of the invention, nucleic acid may be isolated from the multicellular organisms once they are selected and sorted based on their phenotypic characteristics. Methods of isolating DNA or RNA from multicellular organisms and determining its sequence are well known in the art (Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd Ed., 1989; Miller & Calos, eds., *Gene Transfer vectors for Mammalian Cells* , 1987; Ausubel et al., eds., *Current Protocols in Molecular Biology*, 1987; each of which is incorporated herein by reference). The unique nucleic acid identifier can be identified based on its unique nucleotide sequence (e.g., a unique DNA, RNA, or DNA/RNA hybrid molecule sequence). Alternatively, the unique nucleic acid sequence can be identified based on its length. For example, a collection of unique sequences may be, for example, poly-guanine sequences having different lengths. Different length nucleic acid sequences can be detected, e.g., by polyacrylamide gel electrophoresis, or sequencing methods well known in the art (see Sambrook et al. or Ausubel et al., *supra*). Unique nucleotide sequences have been used to mark or "bar code" strains of yeast and such systems are described by Cormack et al. (*Science* (1999) 285:578-582) and Winzeler et al. (*Science* (1999) 285:901-906). The present invention, utilizes unique nucleic acid identifiers in combination with an automated detection and sorting system for multicellular organisms.

In certain preferred embodiments, the unique nucleic acid identifier sequence is amplified. For amplification purposes, the unique nucleotide sequence may be flanked by upstream (5') and downstream (3') universal primer binding sites for universal primers. Universal primers are "universal" in the sense that they are present on each unique nucleotide sequence in each strain of multicellular organism. More particularly, portions of the unique nucleotide sequence at the upstream (5') and downstream (3') regions can be used as primer binding sites. Unique primer binding sites that are separate from the unique nucleic acid identifier may also flank the unique nucleic acid identifier sequence.

It may be preferred that one or more of the primers for amplifying the unique nucleic acid identifier sequence is labeled by a method standard in the art (see, e.g., see Sambrook et al. or Ausubel et al., *supra*). According to this particular embodiment, the labeled primer is incorporated onto the amplified nucleic acid during amplification so that the amplified nucleic acid can be readily detected. In one preferred embodiment, the labeled nucleic acid is hybridized to an array of nucleic acids (i.e., an array of nucleic acid probes) that are complementary to the unique nucleic acid identifiers representative of each genotype present in the mixed population of multicellular organisms (see below). Since the position of each complementary probe sequence on the array is known, the identity of the unique nucleic acid identifier is determined by the location of hybridized complementary probe sequence on the array. Identification of the unique nucleic acid identifiers represented in the sorted population identifies the specific strains present in the sorted population, which share a particular phenotypic characteristic.

In related embodiments, the invention provides a method of selecting and identifying genetically engineered multicellular organisms from a mixed population of genetically engineered multicellular organisms by 1) providing a mixed population of genetically engineered multicellular organisms, wherein the mixed population includes different strains having different promoters that drive expression of a reporter gene, wherein the genotype of each different strain is represented by a different unique identifier, wherein the unique identifier is a unique nucleotide sequence; 2) analyzing the arrangement of spatially distinct, optically detectable, phenotypic characteristics on the multicellular organism, wherein the instrument is

capable of orienting the organism along its longitudinal axis and sorting multicellular organisms having a particular phenotypic characteristic; 3) isolating nucleic acid from the sorted multicellular organisms; 4) optionally amplifying the nucleic acid from the sorted multicellular organisms; 5) contacting the amplified nucleic acid with an array of complementary nucleic acids and allowing the amplified nucleic acid to hybridize to the array; 6) identifying the genotype of the sorted multicellular organism by detecting the location of the amplified nucleic acid hybridized to the array.

Other particularly preferred unique identifiers include unique identifiers that generate a detectable signal. According to the present invention, the detectable signal generated by the unique identifiers is different than any detectable signal generated by the spatially distinct optically detectable phenotypic characteristics on the multicellular organism. Fluorescent signals, e.g., fluorescent moieties or tags, may be used as unique identifiers that generate detectable signals for identification of the multicellular organisms. In certain preferred embodiments, fluorescent moieties or tags are attached directly to the multicellular organisms. For example, a fluorescent moiety or tag may have an attached activated carbonyl, which binds to lysine residues on the surface of an organism, thereby staining the organism with a particular color. Fluorescent moieties or tags of interest include coumarin and its derivatives, e.g. 7-amino-4-methylcoumarin, aminocoumarin, bodipy dyes, such as Bodipy FL, cascade blue, fluorescein and its derivatives, e.g. fluorescein isothiocyanate, Oregon green, rhodamine dyes, e.g. Texas red, tetramethylrhodamine, eosins and erythrosins, cyanine dyes, e.g. Cy3 and Cy5, macrocyclic chelates of lanthanide ions, e.g. quantum dye™, fluorescent energy transfer dyes, such as thiazole orange-ethidium heterodimer, TOTAB, etc. Representative fluorescence detection devices include the Affymetrix GeneArray Scanner (Affymetrix, Santa Clara, Calif.) and Axon GenePix 4000™ microarray scanner (Axon Instruments, Foster City, Calif.). Such fluorescence detection devices may be attached to the automated detection and sorting instrument. Also of interest are nanometer sized particle labels detectable by light scattering, e.g. "quantum dots" (see, e.g., Mattoussi, H, Self-Assembly of CdSe-ZnS Quantum Dot Bioconjugates Using an Engineered Recombinant Protein, *J. Am. Chem. Soc.*, (2000) 122 (49):12142; Empedocles SA et al., Three Dimensional Orientation Measurements of Symmetric Single Chromophores using Polarization

Microscopy, *Nature* (1999) 399:126; Bruchez, MP et al., Semiconductor Nanocrystals as Fluorescent Biological Labels, *Science* (1998) 281:2013), and U.S. Patent No. 5,260,957, each incorporated by reference herein.

According to this aspect of the invention, the present invention provides systems and methods for sorting and identifying genetically engineered multicellular organisms from a mixed population of genetically engineered multicellular organisms by 1) providing a population of genetically engineered multicellular organisms, wherein the genotype of each organism in the population is represented by a different detectable signal, e.g., a fluorescent signal; 2) analyzing the arrangement of spatially distinct, optically detectable, phenotypic characteristics on the multicellular organism, wherein the instrument is capable of orienting the organism along its longitudinal axis and sorting multicellular organisms having a particular phenotypic characteristic; and 3) identifying the genotype of the sorted multicellular organism by detecting the detectable signal with the instrument.

Another preferred unique identifier that generates a detectable signal is an enzyme, such as an enzyme, encoded by a reporter gene (e.g., beta-galactosidase, beta-glucoronidase and the like). Also preferred are proteins that generate fluorescent signals (e.g., green fluorescent protein, yellow fluorescent protein, or red fluorescent protein (see, e.g., Harpur et al. *Nat. Biotechnol.* 19(2):167-169 (2001); Mizuno et al. *Biochemistry* 40(8): 2502-2510 (2001); Huang et al. *Traffic* 2(5):345-357 (2001), each of which is incorporated herein by reference). Preferably, the enzyme generates a different detectable signal than any of the signals generated by the spatially distinct, optically detectable phenotypic characteristics used for phenotypic selection. In addition, in order to utilize this system for a large population, a different enzyme is required for each strain employed in the phenotypic selection. Preferably, the enzyme is detected by the automated instrument simultaneously with detection of the first and second features that constitute the spatially distinct optically detectable phenotypic characteristics. Alternatively, the detectable signal generated by the enzyme may be detected subsequently to the detection of the spatially distinct optically detectable phenotypic characteristics. Detection of the enzyme identifies the genotype of the strain.

Detectable signals also include members of a "signal producing" or "binding partner" system that act in concert with one or more additional members of the same system to provide a detectable signal, preferably on the surface of the multicellular organism. Illustrative of such detectable signals are members of a specific binding pair, such as ligands, e.g. biotin, fluorescein, digoxigenin, antigen, polyvalent cations, chelator groups and the like, where the members specifically bind to additional members of the signal producing system, where the additional members provide a detectable signal either directly or indirectly, e.g. a lectin or antibody conjugated to a fluorescent moiety or an enzymatic moiety capable of converting a substrate to a chromogenic product, e.g., alkaline phosphatase conjugate antibody; and the like.

Other detectable signal producing systems utilize a sequence of six histidines, or alternatively a sequence of alternating histidines and aspartic acids, that allow binding to a nickel or antibody column, wherein the antibody or nickel is bound to one or more of the detectable signals described above. Sequences encoding high affinity epitopes may be employed, such as the FLAG epitope DYKDDDDK (SEQ ID NO: 1), the T7 tag sequence MASMTGGQMG (SEQ ID NO: 2), the S-tag sequence KETAAAKFERQHMDS (SEQ ID NO: 3),or any other sequence that confers high affinity binding to its correlative binding member or a protein reagent. Fusion proteins include glutathione-S-transferase, luciferase, ligands to cell surface receptors, and the like. Such fusions are usually joined via a linker sequence of 3-50 amino acids that promotes the bi-functionality of the protein.

Any of the above molecules that are part of a signal producing system can be linked to the antibodies via cleavable arms (protease sites) or other means. As described herein, antibodies may be linked to detectable signals such as radioactive isotopes or chemically linked to fluorophores or chemiluminescent molecules. Chemical linkage may involve biotinylation using the activated carboxylic acid group or biotin-C11-hydroxysuccinimide ester, which will react with cysteines; any number of other methods generally involving bridging the antibody to a useful chemical moiety, usually accomplished by modifying lysine or other basic residues or through the use of reagents specific for free sulfhydryl groups.

In related embodiments, isotopic labels may also be employed. Isotopic moieties or labels of interest include ^{32}P, ^{33}P, ^{35}S, ^{125}I, and the like. According to

the present invention, such isotopic labels may be attached to the multicellular organism for identification purposes. Alternatively, the isotopic labels may be liked to any of the molecules of the signal producing systems described herein to provide a detectable signal.

Instrument

The present invention uses a fluid flow stream to orient elongate, multicellular, organisms and a narrowly focused, stationary, optical beam to scan them along their major axis as they flow. Features such as cell density, refractility, granularity, and fluorescence can be detected and recorded as a function of position along the length of the oriented organism (*i.e.*, an axial pattern scan). The invention is an improvement in speed and statistical precision over current manual techniques for analyzing multicellular organisms one by one under the microscope. The information from the scan can be used to characterize gene expression and enable physical selection and deposition of phenotypes with desired characteristics, or it can be used to determine alterations in gene expression caused by toxic or therapeutic compounds.

The basic components of the invention are shown in Figure 15. Multicellular organisms are passed serially between a beam generated from a laser (25) and a detector (26). It is an important characteristic of the invention that the organisms are flowing at a constant velocity. For a given set of operating conditions it is possible to determine the amount of time required for an organism to travel from the laser beam to the diverter valve. That time is referred to as the delay. It is adjustable by the operator and typically is between 7 and 9 milliseconds. The passage of the organism through the laser beam causes a signal (27) to be produced, which is analyzed to determine the physical and optical characteristics of the organism including, but not limited to, its size (extinction and time of flight), overall fluorescence characteristics, and the location of fluorescence or light scatter features along the longitudinal axis of the organism. Typically the diverter valve is open so that all of the fluid exiting the flow cell is diverted to waste (28). Depending on the analysis, if the organism displays all of the desired sort parameters, a sort command (29) is generated and the diverter (30) valve is turned off at a time equal to the delay (31) described above. The desired organisms (32), having a desired second feature or other spatially distinct,

optically detectable, phenotypic characteristics, fall directly into a bulk container or microwells (e.g., 24, 96, or 384 microwell plates) (33).

Optical System

The present invention provides an improved instrument that uses light scatter and fluorescence means to optically identify and activate fluidic sorting of multicellular organisms from live populations of organisms such as various life cycle stages of *Caenorhabditis elegans*, the larval stages of *Drosophila melanogaster, and* the embryonic stages of *Danio rerio.* In the case where fluorescence from these organisms is very weak, comparatively high levels of electronic noise accompany the electronic signals that are generated by the fluorescence detector and its associated circuitry. Because these weak signals cannot be used to mark the presence of an organism, another, less noisy, (second) signal must be used to gate fluorescence detection.

Two optical detection methods can be used to generate a low noise "gate" signal. The first method uses a light scatter signal from the organism, which is collected, over an acceptance angle of at least 20 degrees, preferably 20-90 degrees. Light scatter within a collection angle less than 20 degrees is rejected with the use of a spatial filter. Such a light scatter signal unambiguously gates even weak fluorescence signals. These signals can then be correlated with position along the major axis of elongate, multicellular organisms and used as enhanced analysis and sorting parameters. The second optical detection method by which the instrument of the present invention optically identifies and activates fluidic sorting of multicellular organisms from live populations of organisms is by using "extinction." This method utilizes the attenuation of the light entering the detector to signal that a worm has entered the light beam and the return of light entering the detector to signal that a worm has exited the light beam. Instead of light scatter, the attenuation of light with in a collection angle from 0 to 0.5-6.0 degrees in the horizontal axis, and less than 20 degrees in the vertical axis from the laser beam is used as a gating signal to indicate when the elongate multicellular organism is in the beam.

Without limiting the invention, we propose that when an organism intersects the laser beam, light is refracted and diffracted by the organism and thereby the light

is radiated forward in an angular distribution pattern. The light attenuated by the organism at the focus point of the beam in a narrow angle around the axis is referred to as extinguished light, whereas light radiated in a wide angle relative to the optical axis is referred to as scattered light. The detected light distribution pattern may vary depending on the distance of the detector from the beam. If the distance between the organism and the detector is not set appropriately, the detector reads in-between the two regions so that the detector incorrectly measures the presence of an organism. According to the present invention, in the first optical detection method the distance between the organism passing through the beam and the detector is set so that the light scattered by the organism is collected, and light extinguished from the organism is rejected. In the second optical detection method light scattered by the organism is rejected, and light extinguished from the organism is collected.

Fluorescence: Excitation and Emission Optics

Fluorescence is captured at a 90-degree angle to the beam using a fluorescence collection lens. The collected light is transmitted through a series of dichroic mirrors such that different wavelengths are directed to three different photomultiplier detectors. The scattered excitation light, and unwanted fluorescence incident on the photomultiplier tubes is further attenuated using bandpass filters. Any combination of bandpass filters and dichroic mirrors can be used to spectrally resolve the fluorescence signal. However, the standard emission filters are optimized for detection of green, yellow, and red fluorescent proteins and stains (see Figure 14).

Figure 14 illustrates the fluorescence optical excitation and detection optics configuration. The illumination/excitation laser (20) can be coupled to the optical assembly through an optical fiber. The light is then focused to a line within the flow cell (17) with beam formation optics (23). Light scatter measurements are made with a detector and spatial filters positioned in the laser beam path at the optical exit of the flow cell. A collection lens (16) used to collect the fluorescence at a 90-degree angle to the excitation illumination beam, and images the fluorescent light through a series of dichroics mirrors dichroic mirrors (15), and bandpass filters (12, 13, 14) onto a series photomultiplier tube (PMT) detectors. The intersection of the beam, organism, and the collection lenses is called the analysis zone (8).

Gating Using Light Scatter: Optical Illumination/Detection Using Forward Scatter

In the case where fluorescence from these organisms is very weak, comparatively high levels of electronic noise accompany the electronic signals that are generated by the fluorescence detector and its associated circuitry. These weak signals cannot be used to mark the presence of an organism, and another, less noisy, signal must be used to gate fluorescence detection. Axial light loss might be used as such a gate. Another preferred gate could be derived from the low-noise light scatter signal from the organism. Conventional light scatter gating, such as is practiced in flow cytometry of single cells, creates ambiguous signals when used on multicellular organisms and thus leads to false gating of fluorescence. A light scatter detection means and an extinction means is herein described which unambiguously gates these fluorescence signals. These signals can then be correlated with position along the major axis of elongate, multicellular organisms and used as enhanced analysis and sorting parameters.

Traditional optical flow cytometers analyze and sort small particles and single cells in liquid suspension by detecting light scatter within (over) narrow cone or solid angles at various angles to the incident optical beam and fluorescence emission at various wavelengths. Information about cell size and structure can be derived from light scatter collected at different angles. For example, information about size can be derived from light scatter detected at low angles relative to the incident optical beam while information about internal cellular granularity can be derived from light scatter detected at a wide angle (near a right angle) relative to the optical beam. Further, the prior art shows that size of the granular structures to be detected determines the angle and acceptance cone for optimal wide-angle detection.

Light scatter signals collected at specific angles and over narrow cone angles are also used to gate detectors of weak fluorescence from single cells. Weak fluorescence signals cannot be effectively used to mark the presence of a cell in the optical beam because high levels of electronic noise accompany these signals. Noise spikes frequently exceed the threshold level for fluorescence detection and produce false readings that are confused as weakly fluorescing cells. To avoid this, flow

cytometers generally use signals from one or more detectors situated to detect light scatter at one or more angles relative to the beam to produce relatively noise free signals that can effectively discriminate against false fluorescence from electronic noise, and gate true fluorescence from cells (see U.S. Patent 4,284,412, incorporated herein by reference).

One key to the use of these light scatter detectors as fluorescence gates is using a narrow solid angle of detection. For example, so-called "low angle forward scatter" (LAFS) detectors are frequently placed as close as 0.5 degrees to the optical axis and collect light only within a one degree cone. Wide-angle light scatter detectors are frequently placed at positions ranging from approximately 10 degrees to 90 degrees off axis and also collect light within small cone angles of less than five degrees. If the cone angle of collection is not kept as small as possible, then information about granularity and size can become merged. Under these conditions for example, large cells become indistinguishable from small cells and granular cells become indistinguishable from non-granular cells of the same size.

When narrow acceptance cone light scatter (NACLS) detectors are used to monitor the passage of multicellular organisms such as *C. elegans,* three problems arise that do not occur with single cells such as blood cells. First, it is found that the light scatter signal does not necessarily rise above baseline (zero) at the beginning of the passage of the organism through the optical beam, but instead rises at an unpredictably later time. Second, it is found that the light scatter signal does not necessarily return to baseline (zero) at the end of the passage of the organism through the optical beam, but instead returns at an unpredictably early time. Third, it is also found that the light scatter signal frequently returns to baseline (zero) at one or more unpredictable times while the organism is in the beam.

Therefore, the most basic effort to size multicellular organisms based on their "time of flight" through the analysis light beam is thwarted by this unpredictable behavior of light scatter signals that are collected over narrow cone angles. Furthermore, the narrow cone angle light scatter signals that start late are not useful for gating fluorescence signals. Finally, the narrow cone angle light scatter signals that return to baseline early cannot be used to denote the position of fluorescence along the axis of the worm. The signals that return to baseline early can also be

confused with the passage of two or more separate organisms when actually only one passed through the analysis beam.

The present invention does not employ the usual single cell, light scatter detection methods, and instead uses light scatter collection over very wide cone angles when analyzing and sorting multicellular organisms. One aspect of the invention is to collect scattered light over a wide cone angle such as 20 degrees or more. This provides a light scatter signal that becomes positive accurately at the time the organism enters the beam, remains unambiguously above baseline while the organism is in the beam, and returns to baseline accurately at the time the organism exits the beam. This aspect of the invention enables another aspect of the invention, which is to use accurate, unambiguous, light scatter signals collected over wide cone angles to mark the linear position of weak and noisy fluorescence signals along the axis of the organism. The width of the cone angle needed depends upon the type of organism.

In figure 14, and in the preceding paragraphs the fluorescence optical configuration is described. Additionally, a spatial filter and forward scatter detector are placed on the beam axis after the optical exit of the flow cell.

Gating Using Extinction: Optical Illumination/Detection Using Extinction

The present invention also provides a system having two lasers: a red (e.g., 635, 670 nm) diode laser that is used to measure the optical extinction (axial light loss) of the organisms and an excitation laser (e.g., a multine argon-ion laser, etc). Optical extinction of red light provides a highly reliable "triggering" signal that indicates when the worm in passing across the light beam in front of the detector. The signal consistently rises above baseline as the organism enters the beam and does not return to baseline until the organism has left the beam. Use of a longer wavelength (red) light dramatically reduces the anomalous signals observed with using extinction with the argon ion laser beam. Another advantage provided by the red light is that the diode lasers generate far less background fluorescence or "noise," due to higher speed fluctuations in signal intensity compared to the argon ion lasers. This further improves the signal to noise ratio for the optical extinction detection.

In certain preferred embodiments, silicon photodetectors are used to measure the axial light loss signal. Of course, those skilled in the art will appreciate that other types of signal detectors can be used such as avalanche photodiodes, and photomultiplier tubes.

In figure 14, and in the preceding paragraphs, the fluorescence optical configuration is described. Additionally, in the extinction measurement configuration, an illumination laser (e.g. a diode laser) is combined and aligned coaxially with the fluorescence excitation beam (21, 22), the combined beams are then focused into the analysis zone of the flow cell. A long pass filter is placed in the extinction beam path to reject the excitation light from the excitation laser. In addition, an optical slit is placed in between the flow cell and the extinction detector, both located on the optical axis. The function of the slit is to allow the light attenuated by the organism to reach the detector, and rejects the forward scatter light from the organism. The slit has a collection angle of approximately 0.0 to 0.5-6.0 degrees in the horizontal axis and approximately 17 degrees in the vertical axis. The collection angle in the horizontal axis is dependent on the flow cell capillary diameter.

Fluid Mechanics

According to the present invention, the flow cell of the instrument is preferably rectangular with a square cross-section capillary measuring 250 □m on a side for use with *C. elegans*. The flow cell capillary is preferably 1000 □m on a side to accommodate embryos and first through second instar, *D. melanogaster* larvae. However, those skilled in the art will appreciate that a flow cell capillary can have a rectangular flow channel from 250 □m to 500 □m, to 1,000 □m or even 2,000 □m. A 250-□m wide rectangular flow channel is most useful for sorting nematodes such as *Caenorhabditis elegans*. A 500 □m wide rectangular flow channel is most useful for rapid sorting of *Drosophila melanogaster* embryos, although it can also be used to sort smaller model multicelluar organisms such as *C. elegans*. The 1,000-□m wide rectangular flow channel is most useful for sorting *Drosophila melanogaster* up through the second instar larva. Finally, the 2,000 □m wide rectangular flow channel is intended for larger model organisms such as third instar *Drosophila* larva, and zebrafish (*Danio rerio*) embryos and hatchlings. Sheath flow is used to orient these elongate organisms as they emerge from the sample nozzle and enter the flow cell capillary.

This capillary flow cell is located at the line focus of the laser beam. Figure 2 diagrammatically shows the geometric relationship of the flow and the various optical beams. The fluorescent light is collected by simple aspheric lenses or microscope objectives and passed through emission filters to photomultipliers. By virtue of the focused laser beam and the collection lenses, the flowing organism is optically scanned as it passes through the focus. Figure 13 illustrates the instrument of the invention, which includes a rectangular flow channel (1) for through which to pass the organism, particle or cell (2). The sheath (3) flow intersects the rectangular flow channel (1) at the flow convergence zone (4). An aperture to supply an air stream to sort the organism is included (9).

Data Acquisition

There is theoretically no limit to the number of optical features that could be simultaneously monitored. The present invention can measure two, preferably three, more preferably four colors and many different features within each color. However,

without limiting the invention, we propose that the number of signals, e.g., fluorescent features, can vary from 1-3, 1-4 1-10, 10-30, 30-50, or 50-100.

In preferred embodiments, the software on the user interface PC is the only interface between the user and the instrument. The customer is provided, e.g., with a "headless" Linux computer (a computer for which no monitor, keyboard, or mouse is provided and which is intended to prevent user intervention) in addition to the user interface PC. The "headless" PC would not allow user intervention and would in essence become a component of the instrument rather than an independent computer.

One particular profiler that has been used in the inventive system is an option comprised of second, Linux PC and novel software. This software allows for analysis and sorting of animals based on their axial patterns. Axial pattern analysis may be performed on one signal (any color of fluorescence or axial light loss) selectable by the user. The present invention further provides simultaneous axial pattern analysis on two or more signals. There is no theoretical limit to the number of different signals that can be analyzed simultaneously, as described above for light scatter (WACLS) signals.

Signal Analysis

In certain preferred embodiments, the present invention uses the unambiguous light scatter signal from a wide acceptance angle, light scatter (WACLS), or light extinction (EXT) detector as a gate as well as a timing method for the analysis of fluorescence along the axis of the organism. The location of fluorescence along the axis of the organism is an important parameter for analysis and sorting. For example, with *C. elegans*, it is important in many applications to separate males from hermaphrodites. This can be accomplished with a fluorescently labeled lectin (wheat germ agglutinin) that binds to the vulva of the hermaphrodite and the copulatory bursa of the male. These two structures are not easily distinguishable in brightness, but the vulva is located near the midpoint of the organism and the copulatory bursa is located in the tail. Thus, axial location of fluorescence becomes the parameter for differentially analyzing and sorting males and hermaphrodites. This is illustrated schematically in Figure 3 where two oscilloscope traces are shown for single

organisms. One trace (panel A), has a fluorescent peak near the midpoint, and the other trace (panel B) has a fluorescent peak at the tail.

Since there is no fluorescent signal to mark the beginning of the organism in the oscilloscope traces of Figure 3, a means must be established to mark the beginning and end of the passage of the organism through the light beam. This is done by the use of the wide acceptance cone light scatter (WACLS) signal. The start of this signal triggers a clock in the electronic processor that, in turn, causes fluorescent data to be sampled at regular intervals in time while the wide acceptance cone light scatter signal remains above a preset threshold level. Sampling stops when the WACLS signal drops below threshold, denoting the end of the organism.

The following is a parametric representation of a multicellular organism that can be employed through the use of a WACLS signal to gate the sampling of fluorescence along the organism's axis. Consider a WACLS detector that produces signal S 1 and a timing mechanism that samples signals from all other detectors every T microseconds. Assume that there are other light scatter or light absorption detectors situated at various angular positions with respect to the analysis beam. Let the signals from these detectors be denoted by S2, S3, ... Sn. Further assume that there are fluorescence detectors sensitive to various emission wavelengths producing signals F1, F2, F3, ... Fn. The matrix below has columns of data for each detector and rows of data for each sampling interval.

	S1	S2	S3	...	Sn	F1	F2	F3	...	Fn
T1	0	0	0		0	0	0	0		0
T2	a1	0	c1		0	e2	0	0		0
T3	a2	b2	c2		d3	0	f3	g3		0
T4	a3	0	c3		d4	0	f4	0		0
T5	a4	b4	0		0	0	f5	0		0
Tn-1	an-1	0	0		0	0	0	gn-1		0
Tn	0	0	0		0	0	0	0		0

The matrix example above shows a WACLS signal S1 with non-zero entries from time intervals T2 to Tn-1. This is the independent timing signal for all other

detector channels. The other light scatter detectors S2 to Sn are not necessarily WACLS detectors, and therefore have zero values during the time T2 to Tn-1. The fluorescence feature with emission wavelength Fl is small and localized within interval T2. This represents a feature that can be used to mark the "tail" of the organism (see Figure 3).

The fluorescence feature with emission wavelength F2 is not as small (along the axial direction) and occurs at a different location than the F1 feature. The relative location of the feature is established by reference to the timing initiated by the WACLS detector signal S1. In the present embodiment, if the velocity of the organism is known and the "tail" marker is used, then the absolute location of this feature can be determined as well. The fluorescence feature with emission wavelength F3 shows up in two small locations indicated in the WACLS timing sequence as T3 and Tn-1.

Each scanned organism can be represented by a parametric matrix of this kind. While not containing as much information as a microscope image of the organism, the data acquisition times for such matrices are of the order of one microsecond to 250 microseconds, depending on the length of the organism. This high speed is achieved because simple, fast photomultipliers collect the scattered light and no image is formed. In cytometers, images are usually stored by CCD cameras, which are inherently less sensitive than photomultipliers and therefore require more time to collect enough photons to form an image. Imaging times for fluorescence analysis of organisms such as *C. elegans* are of the order of 50 milliseconds, which is from 200 to 10,000 times slower than the time required to collect and store the parametric data described above. In the present embodiment, the sampling time and the speed of the organism determine the spatial resolution of the parametric method. For example, when the organism typically travels at about 500 cm/sec through the analysis beam, then for a five microsecond sampling time the spatial resolution is approximately 25 □m.

Multicellular Organisms

The present invention employs strong fluorescence markers that can be detected against the strong autofluorescence background and used to "bracket" a

section of the signal *(i.e.*, a specific lengthwise region of the organism) where the experimentally created feature is expected to appear and electronically process only this smaller amount of electronic data. This shortened processing task provides valuable processing time for other tasks such as commanding a sorter mechanism before the organism has time to flow beyond the sorter's deflection point. To obtain adequate resolution of axial features the height of the line focus beam must be substantially smaller than the length of the organism analyzed. In addition, the invention provides a means to reduce the variability of the auto fluorescence profile and improve the detection of the markers.

The cells of multicellular organisms like *C. elegans* and *Drosophila melanogaster* (unlike phytoplankton) develop in â reproducible and spatially organized way. This organization is governed by both intracellular and intercellular interactions that provide developing cells with "positional information." Since the spatial location of cells is highly conserved from generation to generation in species such as the fly *D. melanogaster*, and the nematode *C. elegans*, it is possible to express markers that will become stable features of a particular genetic strain of the organism. The present invention takes advantage of this to provide a high signal-to-background "map" of invariant locations along the length of the organism. These locations serve as positional markers to bracket and isolate segments of an axial profile for signal processing.

A strain exhibiting such a "map" (marker pattern) can then be used in a number of research protocols where experimental fluorescence markers are created in a pattern that is independent of the strain marker pattern. The strain marker pattern serves as a reference for the spatial position of the experimentally induced fluorescence markers. Further, the synchronous nature of the markers wherein a marker signal will be found at an expected point allows enhanced detection of the marker signals against background noise.

An example of establishing a fluorescent marker strain of *C. elegans* follows. The genetic manipulations described are well known to those of skill in the art. The invention comprises the use of these genetic constructs. First, one constructs an expression vector that carries a gene for a fluorescent protein, for example ZsYellow from ClonTech, Inc., under control of the *egl-17* promoter sequence. Next, insertion

of this construct into the organism's genome results in expression in the M4 neuron located in the anterior bulb of the pharynx. This expression commences soon after the organism hatches and persists through adulthood. This insertion also results in expression in vulval precursor cells as early as the late L2 larval stage and continuing at lower levels in the vulva of the adult hermaphrodite worm. This pattern of fluorescence expression will be carried as a stable characteristic of the strain resulting from the insertion of the construct into the genome. The fluorescent signal in the head provides the instrument with a means to determine the orientation of the animal while the signal at the vulva provides additional positional information as well as providing some information regarding developmental stage. The invention consists of producing a stable longitudinal pattern and using it as an aid in signal processing. A preferred method is to construct a strain of organism with stable fluorescent markers. The variety of promoters and other genetic constructs that can be employed to achieve this aim is almost limitless.

Generally, the marker fluorescence pattern and the experimentally induced fluorescence pattern will be detectable by different optical channels. That is, if the marker pattern is one of red fluorescence, it is often advantageous to design the experimental treatment (*e.g.*, a screen of potential pharmaceuticals) to show function by producing localized green fluorescence (*i.e.*, non-red fluorescence). In such a scheme, the instrument can be instructed to look for a specific optical pattern using the red fluorescence optics to determine the longitudinal orientation of the organism and to provide additional positional information. Because this signal pattern can be preprogrammed, analysis can be performed more rapidly than if a more complex and variable single color optical system were used. The instrument then compares features in the green fluorescence signal to the positional information in the red. This approach has the further advantage that if the various features of the organism are closely spaced they are more easily resolved if multiple fluorescence markers are used. In some cases a third or even more channels (colors) can be used. Alternatively, it is possible to use only a single optical and electronic channel for both patterns (marker pattern and experimental or test). This would be useful in a case in which a version of the instrument described in patent application number 09/465,215 was employed that utilized only one set of fluorescence optics. It is simply a matter

of balancing instrument complexity and cost against the value of the added information obtained.

The point of the invention is a detectable spatial pattern used for improving signal processing and generally serving as a "map" to pinpoint the location of detectable patterns created or altered by experimental treatments. This does not necessarily require that the genetic-manipulation be used to directly create a fluorescent marker pattern. Exogenous markers such as fluorescently labeled lectins, particles or antibodies can also be used to mark the location of features created by genetic manipulation or of existing structures, such as the vulva, to create a pattern useful for signal processing. That is, the created spatial pattern may not be optically detectable until after treatment with a ligand or with a histochemical process. For example, the promoter or other spatially oriented genetic control element may actually control local expression of an enzyme whose presence is made detectable by a histochemical procedure prior to flow cytometric analysis of the organisms. The detection may be by means of fluorescence or by light absorption or light scatter. Light absorption or scatter may be due to a ligand, a histochemically synthesized dye or compound (*e.g.* precipitation product of a histochemical such as diaminobenzidine or a tetrazolium salt). Also, a particularly dense deposit of a protein or other biomolecule or structure resulting from the genetic manipulation may also be detectable by light scatter or other optical methods. In some cases there may be a useful "inherent" or "latent" pattern within a strain of test organisms. In that case treatment with a lectin or antibody is all that is needed to make the pattern usable.

Since not all markers can be made arbitrarily strong, a means to reduce the effects of autofluorescence is also important. Organisms are not oriented in an azimuthal direction in this invention, but are oriented only along the axis of flow. Consequently, different cellular masses are stimulated into autofluorescence depending upon the azimuthal orientation with respect to the laser axis. In other words, there will be differences in the auto fluorescence profile for each organism that passes through the laser because each organism will be in a different azimuthal orientation (*e.g.*, vulva toward laser or vulva turned away from laser). To compensate for this, a second wavelength band of autofluorescence that lies outside the experimentally created fluorescence band can be monitored, and subtracted from the

total profile. Azimuthal variations in autofluorescence in the two different bands will correlate. Subtraction of the second wavelength band of autofluorescence decreases autofluorescence without significantly altering the measured fluorescence signal from the experimentally created marker. Subtraction reduces the variability in the autofluorescence profile from organism to organism.

Signal processing electronics can be configured to integrate fluorescence signals or to detect the peak of such signals. Integration is useful in reducing electronic noise or laser noise for a spatially diffuse feature, and peak detection is useful in pinpointing the location of a spatially sharp feature. A marker strain profile can be used to trigger different signal processing methods (*e.g.* integration or peak detection) depending on the nature of the experimentally created feature. For example, a given marker strain might produce five spaced-apart marker features along the length of the organism. These marker features are reasonably strong so that peak detection would work well. However, the experimentally induced marker appears between the third and the fourth marker and is fairly diffuse spatially. Therefore, the system could advantageously be programmed to switch from peak detection to integration after the third marker is detected. This would allow optimal detection of the experimentally induced marker. It is only with the use of the tailored marker pattern strains of the present invention that such switching of signal processing electronics becomes possible.

Creating A Marker Strain

A general approach to creating a marker strain of organisms is to genetically introduce a set of features that are readily detected by a flow cytometer. A simple approach is to produce features that can be directly detected by their fluorescence-for example by introducing a gene for a fluorescent protein. Any detectable pattern can be used, however. Enzyme patterns can be detected by histochemical reactions producing a colored or fluorescent product. Proteins can be overexpressed so as to be optically detectable. Other biological products such as fat globules, crystals, or natural pigments can also serve to form an optically detectable pattern. The pattern could be antigenic and be detected by of antibodies, or the pattern could be carbohydrate-based and detectable by addition of lectins. The lectins and antibodies

can be fluorescent, or can be linked to histochemically detectable molecules or optically detectable structures such as microspheres. Although in most instances it will be necessary to employ genetic manipulation to produce an optimal marker strain, some naturally occurring organisms or strains of organisms have cryptic marker features that can be revealed through the application or antibodies, histochemicals or other such methods.

In the case of genetic manipulation it is advantageous to select a promoter that will result in a desired spatial pattern of expression. An example of such a promoter is the *egl-17* promoter of *C. elegans*. This promoter sequence, when inserted at the 5' end of a gene, will result in expression of the gene product (protein) in the M4 neuron and in vulval precursor cells of the organism. The positions of these cells are well characterized and are invariant in a wild-type (N2) background.

The gene controlled by the chosen promoter should encode a detectable product. An example of such a gene product is a fluorescent protein such as the D*sRed* gene (ClonTech, Inc.). As already mentioned, a large variety of other detection methods are available such as those involving enzymatic or antigenic properties. An advantage of a fluorescent protein is that the organism can be analyzed directly with no need for special incubations or other sample preparation.

Standard molecular genetic techniques are used to clone the promoter DNA sequence, the detectable protein gene sequence, and other DNA sequences required for optimal expression in the organism into an appropriate plasmid vector. For example, the present inventors and their associates have constructed a series of expression vectors in which a synthetic intron has been inserted at the 5' end and the *C. elegans unc-54*, a 3' UTR (untranslated region), has been inserted at the 3' end of each ClonTech Reef Coral Protein gene (*AmCyan*, *ZsGreen*, *ZsYellow*, *DsRed*, *DsRed2*, or *AsRed*). The *egl-17* promoter sequence has been inserted upstream of the 5' synthetic intron in each of the expression vectors resulting in *egl-17* expression plasmid constructs for each fluorescent protein.

The expression plasmid DNA is then inserted into the genome of the host organism. One method used for *C. elegans* entails microinjecting plasmid DNA into the gonad of young adult hermaphrodites and selecting progeny that express the detectable marker. Such animals generally carry the marker DNA as an unstable

extrachromosomal array. Additional steps are required to cause the DNA to become integrated into a chromosome and to select the progeny bearing this integration. This is generally accomplished by mutagenizing the animals to introduce random double stranded breaks in chromosomal DNA. During the DNA repair process extrachromosomal sequences can become incorporated into a chromosome. F_2 progeny that have undergone such an incorporation event can be screened. F_2 homozygotes from such an integration event are identified based on their ability to transfer the marker DNA to 100 per cent of their progeny. It should be noted that other methods, including some that result in integration into a specific site in a chromosome could also be used. The point of the present invention is use of the pattern-marked organism as opposed to creation of such an organism.

Figure 9 shows photomicrographs of transgenic organisms where expression of a fluorescent protein is under the control of the *egl-17* promoter. In this case the construct is *egl-17::ZsYellow*. Figures 9A and 9B show light micrographs of two organisms with Figures 9C and 9D showing the corresponding fluorescence images with the M4 head neuron (H) and vulva (V) marked. Diffuse autofluorescence of the gut is discernible between the head and vulva. Figure 10 shows oscilloscope traces of the optical detector signal from flow cytometric analysis of these organisms. Figures 10A and 10B show oscilloscope traces from representative *egl-17::ZsYellow* expressing *C. elegans*. An extinction signal 30 indicates when the organism enters and exits the laser beam. It should be understood that all references herein to extinction could be replaced by wide-angle forward light scatter or another signal shown to effectively indicate presence of an organism. The upper trace 60 represents yellow fluorescence. The yellow fluorescence signal 60 is indicative of the presence of ZsYellow and marks the head and vulva in the organism. These precisely located points of fluorescence represent a marker pattern as used in the present invention. These results should be compared with the trace (Figure 10C) of a control organism lacking the genetic construct.

In cases where mutagenesis has been employed, it is advantageous to remove extraneous mutations by performing several rounds of mating with wild-type organisms and selecting for homozygotes for the inserted marker. Next, the marker must be transferred to an appropriate background strain for the planned assay by

mating. For example, in a RAS pathway assay for new pharmaceuticals one could perform the screen using a *C. elegans lin-15* mutant that already contains a second or possibly a third detectable marker. In that case the positional marker pattern generated above would be transferred by mating into the *lin-15* strain. If only fluorescent markers were going to be utilized, one can simply mix the different DNAs prior to insertion into the genome thus simultaneously adding all markers into the appropriate strain. The desired background is one that shows an optically detectable response to an active compound. This allows the organisms to be used to screen compound libraries for drug candidates. The marker pattern ensures that the detected signal is positionally correct for the screened activity. That is, it is quite likely that test compounds may have multiple activities that could result in positional changes in the expressed signal and/or anomalous expression. The marker pattern allows the system to discriminate between positional shifts in expression. As explained below, the pattern is especially effective in enabling detection of weak signals resulting from test compounds.

Using Marker Pattern To Detect Suppression of a Disease Model Phenotype

Certain disease model pathways involve the inappropriate activation of gene expression in certain tissues or in the migration of certain cell types during development of the animal (which then results in positional changes in marker expression). One such model involves the *Wnt* signaling pathway in *C. elegans*. Components of this pathway appear to be conserved in other organisms and have been shown to function in the development of several forms of cancer, including breast cancer (Nusse, R., and Varmus, H. E., *Cell* (1982) 31, 99-109; Lejeune et al., *Clin Cancer Res* (1995) 1, 215-22) and colon cancer (Morin et al., *Science* (1997)275, 1787-90; Rubinfeld et al., *Science* (1997) 275, 1790-2). *Wnt* signaling in *C. elegans* is involved in controlling the migration of specific cells (Korswagen et al. *Nature* (2000) *in press*). One example is that the proper migration of the QL neuroblast descendants depends upon the proper expression and function of the *Wnt* pathway genes *mab-5* and *egl-20* (Kenyon, C. *Cell* (1986) 46, 477-487; Salser, S.- J., and Kenyon, C., *Nature* (1992) 355, 255-8; Harris, J. et al. *Development* (1996) 122, 3117-31).

If the QL neuroblasts are marked with a fluorescent protein gene, the distance of these cells, which are normally located in the tail of the animal, from the vulva can be measured using marker pattern organisms. Note that the distance between the vulva signal and the M4 neuron signal in the pharynx allows for precise correction for the size of the animal and thus acts as an internal control. The disease model mutant displays inappropriate migration of these cells, or inappropriate expression of fluorescent protein in other cell types. In such a case, a high throughput drug discovery assay involves exposing the animals to compounds and determining which compounds caused the animals to assume a wild-type fluorescence pattern. The marker pattern enables the analysis to readily detect the shift of the positionally incorrect signal into a normal wild-type position. By allowing the signal analysis to focus on limited regions of the organism and/or by allowing a switching of the analysis mode (*e.g.*, peak detection to integration) within specific (lineal) regions of the organism, the invention also allows the unambiguous detection of weak fluorescence signals.

Using Marker Patterns Recognition To Visualize Weak Signals

In some cases the autofluorescence (intrinsic fluorescence of the organism) signal of an organism is great enough to obscure the signal of a marker. In the case of *C. elegans* PY1089 GFP (Green Fluorescent Protein from *Aequorea victoria*) is expressed in two adjacent neurons in the head of the animal. That fluorescence is visible under the microscope as two areas of more concentrated green fluorescence in a background of diffuse autofluorescence. Current automated analytical instrumentation integrate the total fluorescence signal of the organism and are therefore not sensitive to the brighter region within the autofluorescence. Attempts to resolve this strain from wild type *C. elegans* using such automated systems have been unsuccessful thus far. When the electronic signal from the organism is monitored a clear peak signal is seen at one end of the animal. By synchronizing the signal analysis to a known marker pattern, it is possible to determine the orientation of the animal (*e.g.*, head first) and analyze only the specific GFP signal from the head neurons.

In the case of the animal whose oscilloscope traces are depicted in Figure 11D, the total area under the fluorescence profile is 550 units while the area under the fluorescence peak is only 50 units resulting in a signal to noise ratio of 1:10. If, however, the area under the fluorescence peak is compared to an area of comparable width in the region of the animal with the highest autofluorescence that ratio changes to 2:1. If one then considers that the fluorescence signal from the two nerve cell bodies is 35 units, an assay is able to detect the presence of a third fluorescent cell body or the loss of one of the two fluorescent cell bodies.

A useful marker in this situation is the *egl-17* positional markers described above. With *egl-17*::*ZsYellow* as the positional marker the instrument detects the M4 neuron in the anterior portion of the pharynx and the vulval precursor cells and rapidly determines the orientation of the animal as it passes through the analytical chamber. The software looks for the first green fluorescent peak immediately posterior to the M4 neuron and displays the intensity of only that signal. Results include signals such as 18 (no GFP fluorescence), 35 (GFP fluorescence in only one cell), 50 (fluorescence in two cells), 68 (3 cells), and 86 (4 cells).

Figure 11 shows light and fluorescence photomicrographs of a *C. elegans* PY1089 animal are shown in Figure 11A and Figure 11B, respectively. Oscilloscope traces depicting the optical detector signals generated by two different PY1089 animals are shown in Figure 11C and Figure 11D. In Figure 11B the fluorescence from the head neurons (lower end of imaged organism) is clearly visible and is distinguishable from the overall autofluorescence of the animal. Here the animal is oriented such that the two neurons are aligned one on top of the other and only one slightly diffuse spot is observed. In Figures 11C and 11D the one of the traces 30 shows the extinction signals from the animals while the other trace 50 show the green fluorescence signal. The animal in Figure 11D was longer than the animal in Figure 11C and was probably an adult. As expected, both the peak autofluorescence and the total autofluorescence (the area under the curve) are larger for the larger animal. The total fluorescence from the head neurons is approximately the same for both animals. For Figure 11D the total area under the fluorescence curve is 542 units while the area under the GFP peak is 53. Of the peak area approximately 11.5 units are due to autofluorescence while the remaining 41.5 are due to the GFP fluorescence. These

measurements indicate that discrimination could be made automatically provided that a marker pattern is available to automate decisions concerning the organism's orientation.

Figure 12 is a graphic representation (oscilloscope) of the optical detector signals that would result from an *egl-17::Zs Yellow* construct mated into PY1089. The presence of an organism is determined by an extinction signal 30. A yellow fluorescence signal 60 (ZsYellow) clearly marks the head end of the organism (sharp spike used by the software to determine orientation) and provides several other fluorescence peaks along the length of the organism. A more diffuse green fluorescence signal 50 (GFP) is then integrated to determine the head neurons, which immediately follow the yellow M4 neuron peak.

The various marker patterns provided by the present invention allow the software to determine the orientation of elongate organisms, allow the software to specifically measure the position of treatment dependent signals (by comparison to invariant marker pattern signals), allow the software to alter the mode of signal analysis (*e.g.* peak detection versus signal integration) in a positionally controlled manner, and allow the software to limit detailed data analysis to specific positions along the length of the test organism. From the forgoing description a number of uses of the marker pattern organisms will be apparent to those of skill in the art. One method is to produce a test organism that expresses a marker pattern and also variably displays a detectable signal in response to one or more treatments. Generally a treatment will be exposure of the test organism to one or more test compounds, for example, to select active drug candidates from a synthesis library. However, the treatment may also include one or more environmental or other factors that potentiate or otherwise affect the action of the test compound. After the exposure to the treatment, the test organism is analyzed by a flow cytometer. The marker pattern is detected and the analytic software of the system uses the marker pattern to effectively analyze the signal that represents treatment response. As explained above, such analysis would be impossible or much less efficient without use of the marker pattern. It will be appreciated that a major goal is to select out organisms on the basis of their response to the treatment. This requires that data analysis be completed before the organism passes through the sorting section of the flow cytometer. Therefore, data

analysis time is very brief and the enhanced analysis permitted by the use of marker patterns is often crucial.

The following description is provided to enable any person skilled in the art to make and use the invention and sets forth the best modes contemplated by the inventor of carrying out the invention. Various modifications, however, will remain readily apparent to those skilled in the art from a consideration of the specification or practice of the invention disclosed herein, since the general principles of the present invention have been defined herein specifically to provide optical gating devices and methods for use with an optical analyzer/sorter designed for elongated multicellular organisms and improved data processing of optical signals from elongate multicellular organisms by use of a pattern of markers of spaced apart along the long axis of the organisms. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

Identification Using Array Technology

One feature of the subject invention is the array of nucleic acid probes arranged on a support in a pattern at least according to which strands are complementary to the collection of nucleic acid identifiers that represent the strains of multicellular organisms used in the screening assay (see, e.g., U.S. Patent No. 6,287,768, incorporated herein by reference). The probes of the subject arrays are typically nucleic acids or at least mimetics or analogues of naturally occurring polymeric compounds. Biopolymeric compounds of particular interest are deoxyribonucleic acids, as well as ribonucleic acid derivatives thereof, generated through a variety of processes (usually enzymatic processes) such as reverse transcription, etc., e.g. cDNA amplified from RNA (both single and double stranded), cDNA inserts from cDNA libraries as well as chemical synthesis of oligomeric compounds and the like. In certain preferred embodiments, the probe is an mRNA probe.

The probes of the invention, which are located on the array, are capable of hybridizing to the nucleic acid identifiers of the invention. In certain preferred embodiments, the nucleic acid identifiers are isolated and amplified from a phenotypically selected multicellular organism or other physiological source. The

physiological source may be, including plants and animals, where the physiological sources from multicellular organisms may be derived from particular organs or tissues of the multicellular organism, or from isolated cells derived therefrom. For example, the nucleic acid identifiers may be based on the unique promoters used to generate the different strains of multicellular organisms. As but another example, the nucleic acid identifiers may be unique sequences included in each unique strain of multicellular organism.

In other preferred embodiments, the nucleic acid identifiers are obtained from a sample of DNA or RNA isolated from a selected multicellular organism or other physiological source. In obtaining the sample of DNA or RNA to be analyzed from the physiological source from which it is derived, the physiological source may be subjected to a number of different processing steps, where such processing steps might include tissue homogenization, cell isolation and cytoplasm extraction, nucleic acid extraction and the like, where such processing steps are known to those of skill in the art. Methods of isolating DNA or RNA from cell, tissues, organs, or whole multicellular organisms are known to those of skill in the art and are described in Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd Ed., 1989; Miller & Calos, eds., *Gene Transfer vectors for Mammalian Cells* , 1987; Ausubel et al., eds., *Current Protocols in Molecular Biology*, 1987; each of which is incorporated herein by reference).

In the subject arrays, the probes are preferably stably associated with the surface of a support. By stably associated is meant that the probes maintain their position relative to the support under hybridization and washing conditions. As such, the probes can be non-covalently or covalently stably associated with the support surface. Examples of non-covalent association include non-specific adsorption, specific binding through a specific binding pair member covalently attached to the support surface, and entrapment in a matrix material, e.g. a hydrated or dried separation medium, which presents the probe in a manner sufficient for binding, i.e., hybridization, to occur. Examples of covalent binding include covalent bonds formed between the probe and a functional group present on the surface of the support, e.g. -OH, where the functional group may be naturally occurring or present as a member of an introduced linking group, as described in greater detail below.

As mentioned above, the array is typically present on a substrate. Certain substrates are rigid meaning that the support is solid and does not readily bend, i.e. the support is not flexible. Examples of solid materials, which are not rigid supports with respect to the present invention, include membranes, flexible plastic films, and the like. As such, rigid substrates are sufficient to provide physical support and structure to the probes present thereon under the assay conditions in which the array is employed, particularly under high throughput handling conditions.

The substrates upon which the subject patterns of probes are preferably presented in the subject arrays may take a variety of configurations ranging from simple to complex, depending on the intended use of the array. Thus, the substrate could have an overall slide or plate configuration, such as a rectangular or disc configuration, where an overall rectangular configuration, as found in standard microtiter plates and microscope slides, is preferred. For example, the length of the substrates may be at least about 1 cm and may be as great as 40 cm or more, but usually does not exceed about 30 cm and may often not exceed about 15 cm. The width of substrate may be at least about 1 cm and may be as great as 30 cm, but usually does not exceed 20 cm and often does not exceed 10 cm. The height of the substrate will generally range from 0.01 mm to 10 mm, depending at least in part on the material from which the substrate is fabricated and the thickness of the material required to provide the requisite rigidity.

The materials from which the substrate is fabricated should ideally exhibit a low level of non-specific binding of nucleic acid identifier during hybridization or specific binding events. In many situations, it will also be preferable to employ a material that is transparent to visible and/or UV light. Specific materials of interest include: glass; plastics, e.g. polytetrafluoroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and the like; metals, e.g. gold, platinum, and the like; etc.

The substrate of the subject arrays include at least one surface on which a pattern of probe molecules is present, where the surface may be smooth or substantially planar, or have irregularities, such as depressions or elevations. The surface on which the pattern of UNA probes is presented may be modified with one or more different layers of compounds that serve to modulate the properties of the

surface in a desirable manner. Such modification layers, when present, will generally range in thickness from a monomolecular thickness to about 1 mm, usually from a monomolecular thickness to about 0.1 mm and more usually from a monomolecular thickness to about 0.001 mm. Modification layers of interest include: inorganic and organic layers such as metals, metal oxides, polymers, small organic molecules, and the like. Polymeric layers of interest include layers of: peptides, proteins, polynucleic acids or mimetics thereof, e.g. peptide nucleic acids and the like; polysaccharides, phospholipids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneamines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, and the like, where the polymers may be hetero- or homopolymeric, and may or may not have separate functional moieties attached thereto, e.g. conjugated.

The concentration of the probe positions on the surface of the support is selected to provide for adequate sensitivity of binding events with a nucleic acid identifier, where the concentration will generally range from about 1 to 100, usually from about 5 to 50 and more usually from about 10 to 30 ng/mm^2. As summarized above, the subject arrays include a plurality of different probes, where the number of probes is at least 5, usually at least 10, and may be much higher. In some embodiments, the arrays have at least 20 distinct spots, usually at least about 50 distinct spots, and more usually at least about 100 distinct spots, where the number of spots may be as high as 10,000 or higher, but will usually not exceed about 5,000 distinct spots, and more usually will not exceed about 3,000 distinct spots. The density of the spots on the solid surface in certain embodiments is at least about 5/cm^2 and usually at least about 10/cm^2 but does not exceed about 1000/ cm^2, and usually does not exceed about 500/ cm^2, and more usually does not exceed about 300/ cm^2.

The arrays of the subject invention may be used directly in binding assays, i.e., hybridization assays, using well known technologies, e.g. contacting with nucleic acid identifier in a suitable container, under a coverslip, etc, or may be incorporated into a structure that provides for ease of analysis, high throughput, or other advantages, such as in a biochip format, a multiwell format and the like. For example, the subject arrays could be incorporated into a biochip type device in which one has a substantially rectangular shaped cartridge comprising fluid entry and exit ports and a

space bounded on the top and bottom by substantially planar rectangular surfaces, wherein the array is present on one of the top and bottom surfaces.

Alternatively, the subject arrays may be incorporated into a high throughput or multiwell device, wherein each array is bounded by raised walls in a manner sufficient to form a reaction container wherein the array is the bottom surface of the container. Generally in such devices, the devices include a plurality of reaction chambers, each of which contains the array on the bottom surface of the reaction chamber. By plurality is meant at least 2, usually at least 4, and more usually at least 24, where the number of reaction chambers may be as high as 96 or higher, but will usually not exceed 100. The volume of each reaction chamber may be as small as 10 µl but will usually not exceed 500 µl.

The subject arrays may be prepared as follows. The substrate or support can be fabricated according to known procedures, where the particular means of fabricating the support will necessarily depend on the material from which it is made. For example, with polymeric materials, the support may be injection molded, while for metallic materials, micromachining may be the method of choice. Alternatively, supports such as glass, plastic, or metal sheets can be purchased from a variety of commercial sources and used. The surface of the support may be modified to include one or more surface modification layers, as described above, using standard deposition techniques.

Typically, the next step in the preparation process is to prepare the pattern of probe molecules and then stably associate the probe molecules with the surface of the support. The probe molecules may be synthesized using standard procedures. Simultaneously, the nucleic acid identifiers are isolated from the multicellular organisms. Protocols for isolating nucleic acids are described in: Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd Ed., 1989; Miller & Calos, eds., *Gene Transfer vectors for Mammalian Cells* , 1987; Ausubel et al., eds., *Current Protocols in Molecular Biology*, 1987; each of which is incorporated herein by reference, incorporated herein by reference. Such methods typically involve subjection of the original biological source to one or more of tissue/cell homogenization, nucleic acid extraction, chromatography, centrifugation, affinity binding and the like. Nucleic acid identifier preparation can further include one or

more treatments such as: reverse transcription; nuclease treatment; protease digestion; in vitro transcription; DNA amplification; enzymatic or chemical modification of RNA, such as the introduction of functional moieties, such as biotin, digoxigenin, fluorescent moieties, antigens, chelator groups, chemically active or photoactive groups, etc.; and the like.

The probes may be deposited on the support surface using any convenient means, such as by using an "ink-jet" device, mechanical deposition, pipetting and the like. After deposition of material onto the solid surface, it can be treated in different ways to provide for stable association of the probe, blockage of non-specific binding sites, removal of unbound probe, and the like.

Following stable placement of the pattern of probe molecules on the support surface, the resultant array may be used as is or incorporated into a biochip, multiwell or other device, as describe above, for use in a variety of binding applications.

The subject arrays or devices into which they are incorporated may conveniently be stored following fabrication for use at a later time. Under appropriate conditions, the subject arrays are capable of being stored for at least about 6 months and may be stored for up to one year or longer. The subject arrays are generally stored at temperatures between about –20° C to room temperature, where the arrays are preferably sealed in a plastic container, e.g. a bag, and shielded from light.

The present invention utilizes such arrays to detect and identify unique nucleic acid identifiers, which are obtained from phenotypically selected multicellular organisms. Such detection and identification generally involve the following steps: (a) preparation of a nucleic acid identifier or collection of nucleic acid identifiers; (b) contact of the nucleic acid identifiers with the array under conditions sufficient for the nucleic acid identifiers to bind with corresponding complementary probe, e.g. by hybridization or specific binding; (c) removal of unbound nucleic acid identifier from the array; and (d) detection of bound nucleic acid identifier. Each of these steps will be described in greater detail below.

How the nucleic acid identifiers are prepared will necessarily depend on the specific nature of the multicellular organisms from which they originate. The samples may be ribo- or deoxyribonucleotides, as well as hybridizing analogues or mimetics

thereof. The nucleic acid identifiers will have sufficient complementarity to a probe to provide for the desired level of sequence specific hybridization.

Where the probe is a nucleic acid, the length of the probe will generally range from about 10 to 2000 nucleotides, where oligonucleotide probes will generally range in length from about 15 to 100 nucleotides and polynucleotide probes will generally range in length from about 100 to 1000 nucleotides, where such probes may be single or double stranded, but will usually be single stranded. The nucleic acid probes finding use in the subject methods may be synthesized using known chemical or enzymatic synthesis technologies, cloning procedures or obtained from a natural source.

In certain embodiments, the nucleic acid identifier will be labeled to provide for detection in the identification step. By labeled is meant that the nucleic acid identifier includes a member of a signal producing system and is thus detectable, either directly or through combined action with one or more additional members of a signal producing system. Examples of directly detectable labels include isotopic and fluorescent moieties incorporated into, usually covalently bonded to, a moiety of the nucleic acid identifier, such as a nucleotide monomeric unit, e.g. dNMP of the primer, or a photoactive or chemically active derivative of a detectable label which can be bound to a functional moiety of the probe molecule. Isotopic moieties or labels of interest include ^{32}P, ^{33}P, ^{35}S, ^{125}I, and the like.

The next step in the subject method is to contact the nucleic acid identifier with the array under conditions sufficient for binding between the nucleic acid identifier and the probe of the array. For example, where the nucleic acid identifier and probe are nucleic acids, the nucleic acid identifier will be contacted with the array under conditions sufficient for hybridization to occur between the nucleic acid identifier and probe, where the hybridization conditions will be selected in order to provide for the desired level of hybridization specificity.

Contact of the array and nucleic acid identifier involves contacting the array with an aqueous medium comprising the nucleic acid identifier. Contact may be achieved in a variety of different ways depending on the specific configuration of the array. For example, where the array simply comprises the pattern of probes on the surface of a "plate-like" substrate, contact may be accomplished by simply placing the

array in a container comprising the nucleic acid identifier solution, such as a polyethylene bag, small chamber, and the like. In other embodiments where the array is entrapped in a separation media bounded by two plates, the opportunity exists to deliver the nucleic acid identifier via electrophoretic means. Alternatively, where the array is incorporated into a biochip device having fluid entry and exit ports, the nucleic acid identifier solution can be introduced into the chamber in which the pattern of probe molecules is presented through the entry port, where fluid introduction could be performed manually or with an automated device. In multiwell embodiments, the nucleic acid identifier solution will be introduced in the reaction chamber containing the array, either manually, e.g. with a pipette, or with an automated fluid handling device.

Contact of the nucleic acid identifier solution and the probes will be maintained for a sufficient period of time for binding between the nucleic acid identifier and the probe to occur. Although dependent on the nature of the nucleic acid identifier and probe, contact will generally be maintained for a period of time ranging from about 10 min to 24 hrs, usually from about 30 min to 12 hrs and more usually from about 1 hr to 6 hrs.

Following binding of nucleic acid identifier and probe, the resultant hybridization patterns of labeled nucleic acid identifier may be visualized or detected in a variety of ways, with the particular manner of detection being chosen based on the particular label of the nucleic acid identifier, where representative detection means include, e.g., scintillation counting, autoradiography, fluorescence measurement, colorimetric measurement, light emission measurement and the like.

The method may or may not further include a non-bound label removal step prior to the detection step, depending on the particular label employed on the nucleic acid identifier. For example, in homogenous assay formats a detectable signal is only generated upon specific binding of nucleic acid identifier to probe. As such, in homogenous assay formats, the hybridization pattern may be detected without a non-bound label removal step. In other embodiments, the label employed will generate a signal whether or not the nucleic acid identifier is specifically bound to its probe. In such embodiments, the non-bound labeled nucleic acid identifier is removed from the support surface. One means of removing the non-bound labeled nucleic acid

identifier is to perform the well known technique of washing, where a variety of wash solutions and protocols for their use in removing non-bound label are known to those of skill in the art and may be used. Alternatively, in those situations where the probes are entrapped in a separation medium in a format suitable for application of an electric field to the medium, the opportunity may arise to remove non-bound labeled nucleic acid identifier from the probe by electrophoretic means.

The above assays can be used to simultaneously determine the presence or absence of a particular multicellular organism in a particular collection of phenotypically sorted multicellular organisms. The presence or absence of a particular multicellular organism in a particular collection of phenotypically sorted multicellular organisms can be determined by the presence or absence of a hybridization signal at a particular position on the array, which represents the complementary strand to the nucleic acid identifier. To ensure that an accurate level of hybridization is derived, a positive control nucleic acid identifier that is known to hybridize to at least one complementary strand on the array can also be detected, e.g. using a multiplex approach as described above, to provide for a control signal level in order to calibrate the detected signal.

Also provided are kits for analyzing populations of multicellular organisms, particularly mixed populations of multicellular organisms. Such kits according to the subject invention will at least include an array according to the subject invention, where the array may simply include a pattern of probe molecules on a planar support or be incorporated into a multiwell configuration, biochip configuration, or other configuration. The kits may further include a collection of different strains multicellular organisms, wherein each strain contains a unique nucleic acid identifier that can be detected on an array of complementary nucleic acids. The inventive kits may further include one or more additional reagents for use in the assay to be performed with the array, where such reagents include, nucleic acid identifier generation reagents (e.g. buffers, primers, enzymes, labels and the like; reagents used in the binding step, e.g. hybridization buffers; signal producing system members, e.g. substrates, fluorescent-antibody conjugates, etc.; and the like.

Finally, systems that incorporate the subject arrays, particularly the biochip and multiwell configurations of the subject arrays, are provided. Systems of the subject

invention will generally include the array of probes; a fluid handling device capable of contacting the unique nucleic acid identifier fluid and all reagents with the pattern of probe molecules on the array and delivery and removing wash fluid from the array surface; a reader which is capable of providing identification of the location of positive nucleic acid identifier/probe binding events and the intensity of the signal generated by such binding events; and preferably a computer means which capable of controlling the actions of the various elements of the system, i.e. when the reader is activated, when fluid is introduced and the like.

References cited are incorporated herein by reference as if each reference were individually incorporated herein by reference. The teachings of the references are therefore incorporated in their entirety.

Examples

Example 1: Simultaneous NACLS and Fluorescence From *C. elegans*

A light scatter sensor was placed at various angular positions with respect to the optical axis in the forward scatter direction. The collection cone angle was approximately six degrees (NACLS). A photomultiplier with a 20X-collection lens and a barrier filter optimized for fluorescence from GFP was used on the fluorescence detector. The *C. elegans* that were used for this illustration expressed GFP at two locations in the "head" and nowhere else. The oscilloscope traces for light scatter and fluorescence are shown in Figures 4 and 5.

The traces show the passage of the organism through the line focus laser beam. The lower trace (1) is the light scatter signal and the top trace (2) is the fluorescence signal. The x-axis is time. Figure 4 is typical of a class of light scatter traces observed with a NACLS detector. The detector was placed at a 45-degree forward light scatter angle directly below the laser beam axis (below the horizontal plane in Figure 2) as it emerged from the flow cell. No scattered light from the flow cell structures themselves was incident on the detector. The NACLS signal appears to rise at the proper time. The onset of the NACLS trace and the weak autofluorescence trace from the anterior structures of the nematode coincide. The NACLS signal appears to return to baseline after the fluorescent head passes. Unfortunately, the trace returns to baseline approximately during the middle of the passage of the

nematode as well. This would give the false impression that two organisms had passed rather than one. This NACLS signal demonstrates the need for a new, unambiguous trigger and timing signal.

Figure 5 illustrates another problem associated with improper placement of a light scatter detector for triggering. In this example, the same detector was placed in the horizontal plane of Figure 2, but at an angle of 45 degrees to the forward direction. In this case, stray, scattered light from the capillary was incident on the detector. A baseline restoration circuit was used to zero out this light level. The NACLS trace shows a false return to baseline that is caused by the acceptance cone angle being too small, and in fact the signal becomes negative. The negative signal is caused when stray light from the flow cell is blocked by the nematode to an extent that there is more light blockage than there is light scatter. (This signal could not be used as a trigger or timing signal for two reasons. The first reason is that the detector acceptance cone was too small and the second reason is that stray light on the detector became blocked by the passage of the nematode.)

Example 2: Problems Associated with Optical Extinction Signals as Trigger and Timing Signals

Figure 6 illustrates another problem associated with improper placement of a light scatter detector for triggering. In this case a sensor was place directly on axis and in the laser beam. The object was to measure light blockage (extinction) by the organisms. Light extinction is a possible alternative to the preferred WACLS (light scatter) trigger of the present invention. The test *C. elegans* had a single weak region of fluorescence at a neuronal location in the head located slightly posterior to the tip of the "nose." A 40X objective was used to collect more light since this organism was very weakly fluorescent. The extinction sensor collected light over a two-degree cone. In this case, the extinction trace returns to baseline during the passage of the nematode, and even becomes slightly negative. Therefore, this signal could not be used as a trigger or timing signal.

Example 3: Simultaneous WACLS and Fluorescence from *C. elegans*

A photodetector was placed on the optic axis with a collection cone angle of approximately 30 degrees (WACLS). A mask was placed over the center front of the detector to block any directly transmitted light or stray scattered light from the flow cell capillary. This way, the detector collected light scatter from the organisms over a several times wider cone angle than in the previous examples. The photomultiplier with a 40X collection lens and a barrier filter for green fluorescence protein was used to detect fluorescence since the fluorescence signal was very weak.

Figure 7 shows a WACLS signal on the lower trace and the associated fluorescence signal on the upper trace. Note that the WACLS signal begins and ends at the proper time and does not return to baseline during the passage of the nematode. This was a consistent and systematic observation so long as the acceptance angle was sufficiently wide and light from the illuminating beam or the scatter detector did not collect stray light. The particular *C. elegans* used for this example expressed fluorescence along its entire length with 5 to 6 points along the axis where the expression was locally stronger. Some evidence for these local peaks can be seen in the fluorescence trace. The WACLS signal begins and ends at the proper time and does not return to baseline during the passage of the nematode through the laser beam. There were no exceptions to this observation when over 500 nematodes were analyzed. In the examples of useless trigger signals described above almost half of the signals returned to baseline improperly.

Figure 8 also shows the traces for a *C. elegans* with very weak fluorescent protein expression. There is a low level of autofluorescence throughout the length of the organism and two local regions of weak expression near the tail. The WACLS signal begins and ends at the proper time and does not return to baseline during the passage of the nematode through the laser beam. The fluorescence signal is far too noisy to serve as a self trigger and timing signal, however the onset and end of the WACLS signal is strong and unambiguous, and could be used to time and guide an analysis of the fluorescence trace to the location of the two weak peaks.

Example 4: Detection of a Second Feature on *C. elegans*

Figures 16 and 17 show the signals generated from an individual transgenic *C. elegans* that expresses fluorescent protein (ZsYellow, ClonTech Corp.) in one head

neuron and in the vulva. The trace in Figure 16 is a graphic representation that shows the signal stored by the Profiler instrument. It is this data on which the Profiler sorting algorithms act. The ZsYellow protein is expressed using the *egl*-17 *C. elegans* promoter, which derives expression in one head neuron (left side of trace) and in the vulva (middle peak). This expression serves as marker pattern features to provide orientation and positional information.

Figure 17 is a graphic representation of the same animal used to obtain the data for Figure 16 that shows the un-processed electronic signals generated by the animal as it passed through the laser beam. The lower trace in Figure 17 shows the EXT (Axial Light Loss) signal that acts as a "triggering" parameter for the instrument. The upper trace shows the un-processed yellow fluorescence signal from the animal and is equivalent to the trace in Figure 16. Both the EXT (trace 1) and the yellow fluorescence signal (trace 2) were obtained using an oscilloscope. This animal is an example of a marker organism that carries a stable longitudinal pattern that can be used as an aid in signal processing. In this case, the markers allow the software to distinguish head from tail (head neuron) and provide additional positional information (vulva).

Figure 18 is a photomicrograph of a transgenic *C. elegans* expressing ZsYellow under the control of the *egl-17* promoter with the upper panel showing a white light image of the organism and the corresponding fluorescence image shown in the lower panel. There are two bright spots in the fluorescence image that correspond to ZsYellow fluorescence, one in the middle of the animal (vulva) and one small spot in the head of the animal (right side of the image).

Materials and Methods

The method used to create the organism depicted in the above images is described above. The method used to generate the traces in the first two panels was as follows. *C. elegans* were grown on agar media with a lawn of *Escherichia coli*. The plates were removed from the incubator (20-degrees C) and flooded with a salts buffer. The liquid was gently swirled to lift the animals off the agar surface and the animals were transferred to a conical centrifuge tube. The animals were allowed to settle to the bottom of the tube and the buffer was aspirated to waste. The animals were resuspended in buffer and placed into a sample cup of the instrument at a

concentration less than or equal to 1 per microliter. Sort parameters were set using the COPAS™ Profiler software on the user interface PC and using the Profiler software on the Linux PC. The sample was run by clicking the Acquire button followed by the Manual Sort button.

The following claims are thus to be understood to include what is specifically illustrated and described above, what is conceptually equivalent, what can be obviously substituted and also what incorporates the essential idea of the invention. Those skilled in the art will appreciate that various adaptations and modifications of the just described preferred embodiment could be configured without departing from the scope of the invention. The illustrated embodiment has been set forth only for the purposes of example and that should not be taken as limiting the invention. Therefore, it is to be understood that, within the scope of the appended claims, the invention may be practiced other than as specifically described herein.

Example 5: Phenotypic Analysis and Identification

Transgenic strains of *Caenorhabditis elegans* (or any other multicellular organism, as described herein) are generated and their initial profiles collected to establish the baseline data for all the genes and identify the spectrum of all possible profiles. Initially about 12,000 strains are constructed, increasing to about 20,000 strains. Each transgenic strain expresses the same fluorescent protein off of a different promoter. That is, there are 12,000 to 20,000 promoters that drive expression of a fluorescent protein. Each strain also has a unique nucleic acid identifier sequence, or the possibility of generating a unique sequence by PCR from that strain. In the present Example, 2.5 to 3 kb upstream (5') of the start site of transcription for the mRNA encoding the fluorescent protein is located a unique nucleic acid identifier sequence. As shown in Figure A, the upstream and downstream primers are universal primers and the sequence between them is the unique identifier.

A mixed population of strains that uniformly give one particular phenotype is provided. Starter cultures of the transgenic strains are grown in 96-well plates. ReFLx, an add-on module of the COPAS instrumentation that sequentially aspirates a sample from each well of a multiwell plate and passes the sample to the flow cell for

analysis in an automated fashion, is used to dispense a set number of each starter culture into a pool of starter cultures. The mixed populations of transgenic *C. elegans* are next grown in the presence of a test treatment or under specific conditions. According to the present example, the test treatment is a test compound or drug. The organisms are collected and analyzed with a profiler and those having a particular expression profile are sorted away from the organisms that lack the particular expression profile. Specifically, animals expressing profiles that represent the expression pattern of important biochemical pathways are collected and their nucleic acid identifier determined, as described below. Transgenic strains that inappropriately express a different pattern than their normal pattern of expression so that they are now present in this collection are informative. This indicates that the test treatment changed or altered the expression pattern of a particular gene or gene pathway. Alternatively, the assay initially screens all 12,000 or 20,000 strains and organisms for a particular expression profile (e.g., a particular phenotype such as promoter driven expression in a certain tissue specific location). Transgenic strains that should be in the collection of a sort of a particular pattern and are missing are also informative. This indicates that the test treatment causes a reduction or loss of that particular strain, most likely to changing its pattern of expression.

Once the organisms are sorted, the genotypic identity of the sorted organisms is determined. DNA is prepared from the sorted organisms and amplified by PCR using the universal primers. The amplified DNA is labeled with a detectable radioactive or fluorescent label. The amplified, labeled DNA is then hybridized to a DNA chip that contains the complementary strands of each unique nucleic acid identifier being used to encode the genotypes of the multicellular organisms screened. Since the locations on the chip of the complementary strands is known, the identity of the amplified DNAs, and therefore the identity of the sorted organisms, is determined based on which locations generate a hybridization signal.

The results are divided into four classes as follows: 1) organisms expected to be present; 2) organisms expected to be present, but have altered profiles; 3) organisms not expected to be present, but are present; 4) organisms expected to be present, but are not present.

What is claimed is:

Claims

1. A system for sorting multicellular organisms comprising:

a population of multicellular organisms comprising a plurality of spatially distinct, optically detectable, phenotypic characteristics; and

an instrument for detecting the location of the spatially distinct, optically detectable, phenotypic characteristic on the multicellular organism and for orienting the multicellular organism along its longitudinal axis.

2. The system of claim 1, wherein the spatially distinct, optically detectable, phenotypic characteristics comprise a marker pattern comprising a plurality of spatially consistent first features spaced apart along a length of each organism and at least one second feature modifiable or inducible when the population is subjected to a test treatment.

3. The system of claim 1, wherein the instrument is a flow cytometer equipped to process elongate multicellular organisms.

4. The system of claim 1, wherein the instrument measures a gating signal for detecting the spatially distinct, optically detectable, phenotypic characteristic over background signals.

5. The system of claim 4, wherein the gating signal comprises light scattered in the forward direction.

6. The system of claim 4, wherein the gating signal comprises light attenuated by the organism in the forward direction.

7. The system of claim 1, wherein the instrument further comprises:

a source containing multicellular organisms in a fluid suspension;

means for causing the fluid suspension to move in a direction of flow;

means for aligning the elongate multicellular organisms relative to the direction of flow;

a light source for producing an optical beam through which the elongate multicellular organisms pass after becoming aligned;

a first optical detector for detecting light over a solid angle of at least 20 degrees and over a collection angle of approximately 0.0 to 6.0 degrees in the horizontal axis and approximately 17 degrees in the vertical axis, for detecting passage of said organisms through said optical beam; and

a fluid switch downstream of a point where said organisms pass through said optical beam, said switch responsive to the first optical detector to allow detected objects to pass to a sample container.

8. The system of claim 7, further comprising additional optical detectors for detecting sequential optical characteristics arrayed along a length of the multicellular organism wherein outputs of said detectors are gated by an output of the first optical detector to produce gated outputs.

9. The system of claim 8, further comprising a data representation of the sequential optical characteristics comprised of the outputs of the additional optical detectors.

10. The system of claim 9, further comprising a controller connected to the fluid switch and operative to cause said switch to select multicellular organisms showing data representations meeting predetermined criteria.

11. A method for sorting multicellular organisms comprising the steps of:

providing a population of test organisms, wherein each member of the population displays at least one spatially distinct, optically detectable, phenotypic characteristic;

analyzing the arrangement of spatially distinct, optically detectable, phenotypic characteristics of each population member; and

depositing members of the population based on the arrangement of spatially distinct, optically detectable, phenotypic characteristics.

12. The method of claim 11, wherein the spatially distinct, optically detectable, phenotypic characteristics comprise a marker pattern comprising a plurality of spatially consistent first features spaced apart along a length of each organism and at least one second feature modifiable or inducible when the population is subjected to a test treatment.

13. The method of claim 12, wherein the organisms are selected based on the location of the second feature with respect to the first features along the length of each organism.

15. An instrument for analyzing and selectively dispensing elongate multicellular organisms comprising:

a source containing multicellular organisms in a fluid suspension;

means for causing the fluid suspension to move in a direction of flow;

means for aligning the elongate multicellular organisms relative to the direction of flow;

a light source for producing an optical beam through which the elongate multicellular organisms pass after becoming aligned;

a first optical detector for detecting light over a solid angle of at least 20 degrees and over a collection angle of approximately 0.0 to 6.0 degrees in the horizontal axis and approximately 17 degrees in the vertical axis for detecting passage of said organisms through said optical beam; and

a fluid switch downstream of a point where said organisms pass through said optical beam, said switch responsive to the first optical detector to allow detected objects to pass to a sample container.

16. The instrument of claim 15, further comprising additional optical detectors for detecting sequential optical characteristics arrayed along a length of the multicellular organism wherein outputs of said detectors are gated by an output of the first optical detector to produce gated outputs.

17. The instrument of claim 16, further comprising a data representation of the sequential optical characteristics comprised of the outputs of the additional optical detectors.

18. The instrument of claim 17, further comprising a controller connected to the fluid switch and operative to cause said switch to select multicellular organisms showing data representations meeting predetermined criteria.

19. A method of selectively dispensing elongate multicellular organisms comprising the steps of:

centering and orienting the sample objects in a flowing fluid stream;

passing the fluid stream through a sensing zone;

optically detecting the presence of a multicellular organism passing through the sensing zone by means of a light scatter sensor that has an acceptance angle of at least 20 degrees and over a collection angle of approximately 0.0 to 6.0 degrees in the horizontal axis and approximately 17 degrees in the vertical axis;

creating a data representation of sequential optical characteristics of the multicellular organism comprising output signals from additional optical sensors;

diverting at least some portion of the fluid stream with a switched fluid stream based on the data representation so as to collect ones of the multicellular organisms remaining in portions of the sample stream that were not diverted.

20. The method of claim 19, further comprising the step of exposing the multicellular organisms collected in the step of diverting to a test chemical or test environment.

21. The method of claim 19 further comprising the step of exposing the multicellular organisms to a test chemical or a test environment prior to the detecting step to determine whether the data representation is altered by the test chemical or the test environment.

22. A data structure representative of an oriented elongate multicellular organism containing indicia of sequential optical characteristics disposed along a length of said organism, said data structure comprised of stored sequential outputs derived from optical sensors arranged to receive optical energy emanating from the elongate multicellular organism as said organism passes through an optical beam wherein a signal from a light scatter sensor that has an acceptance angle of at least 20 degrees and over a collection angle of approximately 0.0 to 6.0 degrees in the horizontal axis and approximately 17 degrees in the vertical axis is used to create or utilize the data structure.

23. A process for analyzing elongate multicellular organisms by flow cytometry comprising the steps of:

creating a population of test organisms wherein each member of the population displays a marker pattern, said marker pattern representing a plurality of spatially consistent first features spaced apart along a length of each organism and wherein each member of the population also displays at least one of a second feature modifiable or inducible when the population is subjected to a test treatment, each of said first and said second features being detectable through analysis with a flow cytometer;

subjecting the population to a test treatment;

analyzing members of the population with a flow cytometer equipped to process elongate multicellular organisms; detecting the marker pattern on the members analyzed; and

using the detected marker pattern to determine status of the second feature on each of the members analyzed.

24. The process according to claim 23, wherein the step of creating a population includes the step of producing a transgenic organism.

25. The process according to claim 24, wherein the step of producing a transgenic organism includes choice of a particular promoter.

26. The process according to claim 23, wherein the marker pattern is detectable by a flow cytometer by use of detection means selected from the group consisting of light scatter, light absorption and fluorescence.

27. The process according to claim 23, wherein the step of subjecting the population to a test treatment includes contacting the population with a candidate drug molecule.

28. The process according to claim 23, wherein the second feature responds to the test treatment by a change detected as an optical signal, the change being one selected from the group consisting of an increased signal, a decreased signal or a positionally altered signal.

29. The process according to claim 23, wherein the step of using the detected marker pattern includes the step of determining a longitudinal orientation of each member of the population analyzed.

30. The process according to claim 23, wherein the step of using the detected marker pattern includes the step of limiting analysis of data corresponding to the second feature to a particular longitudinal region of each of the members analyzed.

31. The process according to claim 23, wherein the step of using the detected marker pattern includes the step of altering a mode data analysis for data

corresponding to the second feature in a particular longitudinal region of each of the members analyzed.

32. The process according to claim 31, wherein the mode of data analysis is selected from the group consisting of signal peak analysis and signal integration.

33. The process of claim 23, wherein the step of analyzing members of the population with a flow cytometer comprises selecting a gating signal.

34. The process of claim 33, wherein the gating signal comprises light scattered in the forward direction.

35. The process of claim 33, wherein the gating signal comprises light attenuated by the organism in the forward direction.

36. A process for preparing a model strain of elongate multicellular organisms intended for specialized flow cytometry analysis comprising the steps of:

creating a marker strain of organisms wherein each member of the strain displays a marker pattern, said marker pattern representing a plurality of marker features spaced apart along a length of each organism and spatially consistent from member to member, said marker features being detectable through analysis with a flow cytometer;

creating a test strain of organisms wherein each organism of the test strain displays at least one test feature modifiable or inducible when the test strain is subjected to a test treatment, said test features being detectable through analysis with a flow cytometer; and

creating a model strain by combining the marker pattern from the marker strain with the test features from the test strain so that each organism of the model strain displays both the marker pattern and at least one test feature.

37. An organism belonging to a model strain produced by the process of claim 36.

38. A process for analyzing elongate multicellular organisms by flow cytometry comprising the steps of:

subjecting a population of the model strain of claim 36 to a test treatment;

analyzing members of the subjected population with a flow cytometer equipped to process elongate multicellular organisms;

detecting the marker pattern on the members analyzed; and

using the detected marker pattern to determine status of the test feature on each of the members analyzed.

39. The process according to claim 36, wherein the step of creating a population includes the step of producing a transgenic organism.

40. The process according to claim 36, wherein the step of producing a transgenic organism includes choice of a particular promoter.

41. The process according to claim 38, wherein the marker pattern is detectable by a flow cytometer by use of detection means selected from the group consisting of light scatter, light absorption and fluorescence.

42. The process according to claim 38, wherein the step of subjecting the population
to a test treatment includes contacting the population with a candidate drug molecule.

43. The process according to claim 38, wherein the test feature responds to the test treatment by a change detected as an optical signal, the change being one selected from
the group consisting of an increased signal, a decreased signal or a positionally altered signal.

44. The process according to claim 38, wherein the step of using the detected marker
pattern includes the step of determining a longitudinal orientation of each member of the
population analyzed.

45. The process according to claim 38, wherein the step of using the detected marker
pattern includes the step of limiting analysis of data corresponding to the second feature
to a particular longitudinal region of each of the members analyzed.

46. The process according to claim 38, wherein the step of using the detected marker
pattern includes the step of altering a mode data analysis for data corresponding to the

second feature in a particular longitudinal region of each of the members analyzed.

47. The process according to claim 46, wherein the mode of data analysis is selected from the group consisting of signal peak analysis and signal integration.

48. A system for identifying genetically engineered multicellular organisms having a particular expression profile from a population of genetically engineered multicellular organisms, comprising:

a population of genetically engineered multicellular organisms comprising a plurality of spatially distinct, optically detectable, phenotypic characteristics;

an instrument for detecting the location of a spatially distinct, optically detectable, phenotypic characteristic on the multicellular organism and for orienting the multicellular organism along its longitudinal axis, wherein the instrument sorts multicellular organisms having a particular phenotypic characteristic; and

an identification system for identifying the genotype of each sorted multicellular organism having the particular phenotypic characteristic.

49. The system of claim 48, wherein the population is a mixed population of genetically engineered multicellular organisms.

50. The system of claim 48, wherein the multicellular organisms are 25 to 1,250 micrometers in diameter.

51. The system of claim 48, wherein the multicellular organism is selected from the group consisting of *Caenorhabditis elegans* of any developmental stage, *Drosophila melanogaster* eggs, embryos, and larva; and *Danio rerio* eggs, embryos, and larva.

52. The system of claim 48, wherein the spatially distinct, optically detectable, phenotypic characteristics comprise a marker pattern comprising a plurality of

spatially consistent first features spaced apart along a length of each organism and at least one second feature modifiable or inducible when the population is subjected to a test treatment.

53. The system of claim 48, wherein the instrument is a flow cytometer equipped to process elongate multicellular organisms.

54. The system of claim 48, wherein the identification system comprises a unique identifier for each genetically engineered multicellular organisms in the mixed population having a different genotype.

55. The system of claim 54, wherein the unique identifier comprises a unique nucleic acid sequence.

56. The system of claim 55, wherein the unique nucleic acid sequence is flanked by upstream and downstream universal primer binding sites that can be used to amplify the unique nucleic acid sequence to generate an amplified unique nucleic acid sequence.

57. The system of claim 56, wherein the amplified unique nucleic acid sequence is labeled.

58. The system of claim 56 or 57, wherein the amplified unique nucleic acid sequence is hybridized to a complementary sequence on a nucleic acid array.

59. The system of claim 58, wherein the location of the hybridized sequence on the array identifies the genotype of the sorted multicellular organism.

60. The system of claim 54, wherein the unique identifier comprises a detectable signal.

61. The system of claim 60, wherein the detectable signal is detectable by the instrument so that detection and identification occur simultaneously.

62. The system of claim 60, wherein the identity of the detectable signal identifies the genotype of the sorted multicellular organism.

63. The system of claim 60, 61, or 62, wherein the detectable signal comprises a reporter gene expressed by the multicellular organism, wherein the reporter gene generates a different detectable signal than any of the spatially distinct, optically detectable, phenotypic characteristics.

64. The system of claim 60, 61, or 62, wherein the detectable signal is a colorimetric signal.

65. The system of claim 64, wherein the colorimetric signal is a fluorescent signal.

66. The system of claim 65, wherein the fluorescent signal is on the multicellular organism.

67. The system of claim 60, 61, or 62, wherein the detectable signal comprises a chemical signal.

68. The system of claim60, 61, or 62, wherein the detectable signal comprises a radioactive signal.

69. The system of claim 60, 61, or 62, wherein the detectable signal comprises a polypeptide signal.

70. A method for sorting genetically engineered multicellular organisms from a population of genetically engineered multicellular organisms comprising the steps of:

providing a population of test organisms, wherein each member of the population displays a plurality of spatially distinct, optically detectable, phenotypic characteristic;

analyzing the arrangement of spatially distinct, optically detectable, phenotypic characteristics of each population member using an instrument capable of detecting the location of a spatially distinct, optically detectable, phenotypic characteristic on the multicellular organism, wherein the instrument is capable of orienting the multicellular organism along its longitudinal axis and sorting multicellular organisms having a particular phenotypic characteristic; and

identifying the genotype of the sorted members of the population.

71. The method of claim 70, wherein the population of genetically engineered multicellular organisms is a mixed population of multicellular organisms.

72. The method of claim 71, wherein the multicellular organisms are 25 to 1,250 micrometers in diameter.

73. The method of claim 71, wherein the multicellular organism is selected from the group consisting of *Caenorhabditis elegans* of any developmental stage, *Drosophila melanogaster* eggs, embryos, and larva; and *Danio rerio* eggs, embryos, and larva..

74. The method of claim 71, wherein the spatially distinct, optically detectable, phenotypic characteristics comprise a marker pattern comprising a plurality of spatially consistent first features spaced apart along a length of each organism and at least one second feature modifiable or inducible when the population is subjected to a test treatment.

75. The method of claim 71, wherein the step of analyzing comprises passing each multicellular organism through an instrument for detecting the location of the spatially distinct, optically detectable, phenotypic characteristics on the multicellular organism and for orienting the multicellular organism along its longitudinal axis.

76. The method of claim 71, wherein the organisms are sorted based on the location of the second feature with respect to the first features along the length of each organism.

77. The method of claim 71, wherein the organisms are sorted based on the level of expression of the second feature.

78. The method of claim 71, wherein the organisms are sorted based on the timing of expression of the second feature.

79. The method of claim 71, wherein the step of identifying comprises detecting a unique identifier on each multicellular organism.

80. The method of claim 79, wherein the unique identifier comprises a unique nucleic acid sequence.

81. The method of claim 80, wherein the unique nucleic acid sequence is flanked by upstream and downstream universal primer binding sites that can be used to amplify the unique nucleic acid sequence to generate an amplified unique nucleic acid sequence.

82. The method of claim 81, wherein the amplified unique nucleic acid sequence is labeled.

83. The method of claim 81 or 82, wherein the amplified unique nucleic acid sequence can be hybridized to a complementary sequence on a nucleic acid array.

84. The method of claim 83, wherein the location of the hybridized sequence on the array identifies the genotype of the sorted multicellular organism.

85. The method of claim 79, wherein the unique identifier comprises a detectable signal.

86. The method of claim 85, wherein the detectable signal is detectable by the 87. The method of claim 85, wherein the identity of the detectable signal identifies the genotype of the sorted multicellular organism.

88. The system of claim 85, 86, or 87, wherein the detectable signal comprises a reporter gene expressed by the multicellular organism, wherein the reporter gene generates a different detectable signal than any of the spatially distinct, optically detectable, phenotypic characteristics.

89. The system of claim 85, 86, or 87, wherein the detectable signal is a colorimetric signal.

90. The system of claim 89, wherein the colorimetric signal is a fluorescent signal.

91. The system of claim 90, wherein the fluorescent signal is on the multicellular organism.

92. The system of claim 85, 86, or 87 wherein the detectable signal comprises a chemical signal.

93. The system of claim 85, 86, or 87, wherein the detectable signal comprises a radioactive signal.

94. The system of claim 85, 86, or 87, wherein the detectable signal comprises a polypeptide signal.

95. A process for analyzing elongate multicellular organisms by flow cytometry comprising the steps of:

subjecting a population of genetically engineered multicellular organisms to a test treatment, wherein each member of the population displays at least one test feature, wherein the test feature is modifiable or inducible when the genetically engineered multicellular organism is subjected to the test treatment;

analyzing members of the subjected population with a flow cytometer equipped to detect changes in the test feature due to the test treatment, wherein the flow cytometer is equipped to sort multicellular organisms having a particular change in the test feature; and

identifying the genotype of the sorted members of the population.

96. The process of claim 95, wherein the population of genetically engineered multicellular organisms is a mixed population of genetically engineered multicellular organisms.

97. The process of claim 95, wherein the multicellular organisms are 25 to 1,250 micrometers in diameter.

98. The system of claim 95, wherein the multicellular organism is selected from the group consisting of *Caenorhabditis elegans* of any developmental stage, *Drosophila melanogaster* eggs, embryos, and larva; and *Danio rerio* eggs, embryos, and larva.

99. The process of claim 95, wherein the test feature comprises a detectable enzyme expressed by a particular promoter.

100. The process of claim 95, wherein the step of subjecting the population to a test treatment includes contacting the population with a candidate drug molecule.

101. The process of claim 95, wherein the test feature responds to the test treatment by a change detected as an optical signal, the change being one selected from the group consisting of an increased signal, a decreased signal, a positionally altered signal, or a temporally altered signal.

102. The process of claim 95, wherein the step of identifying comprises detecting a unique identifier on each multicellular organism.

103. The process of claim 102, wherein the unique identifier comprises a unique nucleic acid sequence.

104. The process of claim 103, wherein the unique nucleic acid sequence is flanked by upstream and downstream universal primer binding sites that can be used to amplify the unique nucleic acid sequence to generate an amplified unique nucleic acid sequence.

105. The process of claim 104, wherein the amplified unique nucleic acid sequence is labeled.

106. The process of claim 104 or 105, wherein the amplified unique nucleic acid sequence can be hybridized to a complementary sequence on a nucleic acid array.

107. The process of claim 106, wherein the location of the hybridized sequence on the array identifies the genotype of the sorted multicellular organism

108. The process of claim 102, wherein the unique identifier comprises a detectable signal.

109. The process of claim 108, wherein the detectable signal is detectable by the instrument so that detection and identification occur simultaneously.

110. The process of claim 108, wherein the identity of the detectable signal identifies the genotype of the sorted multicellular organism.

111. The process of claim 108, 109, 110, wherein the detectable signal comprises a reporter gene expressed by the multicellular organism.

112. The system of claim 108, 109, 110, wherein the detectable signal is a colorimetric signal.

113. The system of claim 112, wherein the colorimetric signal is a fluorescent signal.

114. The process of claim 113, wherein the fluorescent tag is on the multicellular organism.

115. The system of claim 108, 109, 110, wherein the detectable signal comprises a chemical signal.

116. The system of claim 108, 109, 110, wherein the detectable signal comprises a radioactive signal.

117. The system of claim 108, 109, 110, wherein the detectable signal comprises a polypeptide signal.

118. A system for identifying elongate objects having a particular characteristic from a population of elongate objects, comprising:

a population of elongate objects comprising a plurality of spatially distinct, optically detectable, characteristics;

an instrument for detecting the location of a spatially distinct, optically detectable, characteristic on the elongate objects and for orienting the elongate objects along their longitudinal axis, wherein the instrument sorts elongate objects having a particular characteristic; and

an identification system for identifying the elongate objects having the particular characteristic.

119. The system of claim 118, wherein the elongate object is between 25 micrometers and 1,250 micrometers .

120. A method for sorting elongate objects from a population of elongate objects comprising the steps of:

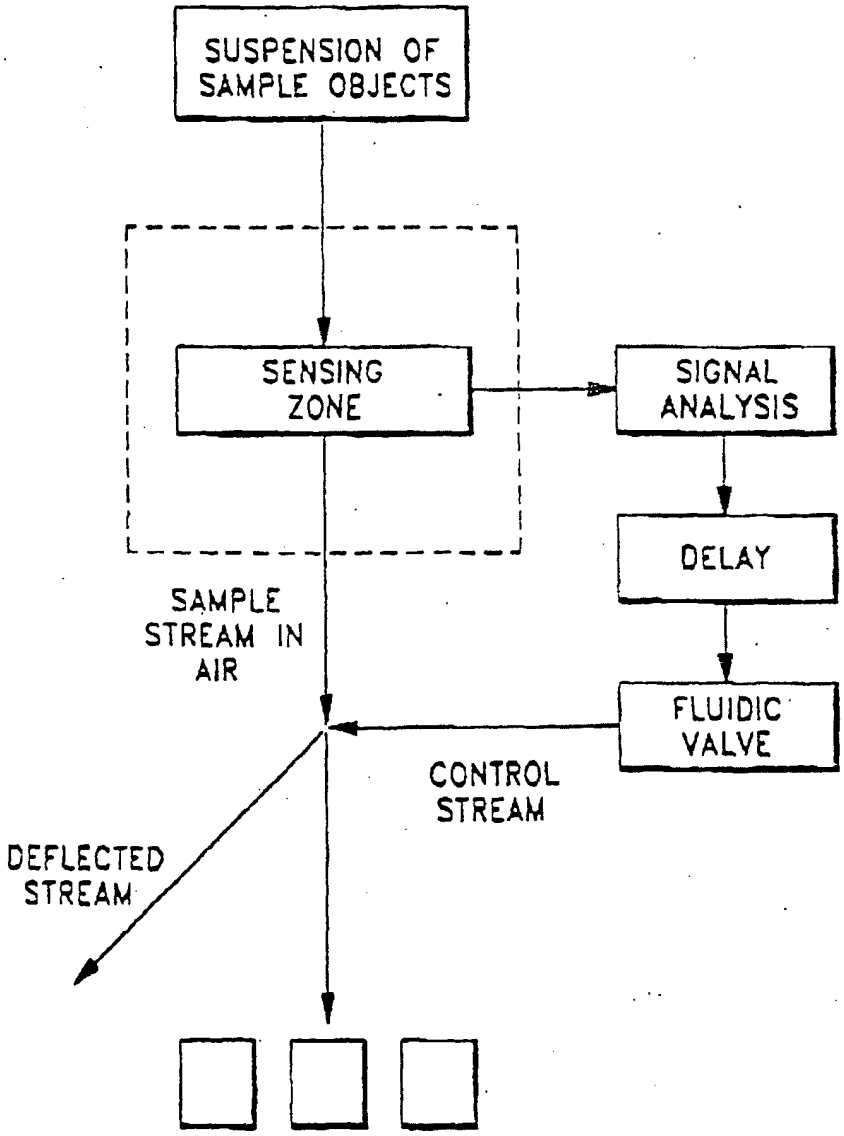
providing a population of elongate objects, wherein each member of the population displays a plurality of spatially distinct, optically detectable, characteristic;

analyzing the arrangement of spatially distinct, optically detectable, characteristics of each population member using an instrument capable of detecting the location of a spatially distinct, optically detectable, characteristic on the elongate object, wherein the instrument is capable of orienting the elongate object along its longitudinal axis and sorting elongate objects having a particular characteristic; and

identifying the elongate objects of the population.

121. The method of claim 120, wherein the elongate object is 25 to 1,250 in diameter.

FIGURE 1


SUSPENSION OF SAMPLE OBJECTS
SENSING ZONE
SIGNAL ANALYSIS
DELAY
SAMPLE STREAM IN AIR
FLUIDIC VALVE
CONTROL STREAM
DEFLECTED STREAM

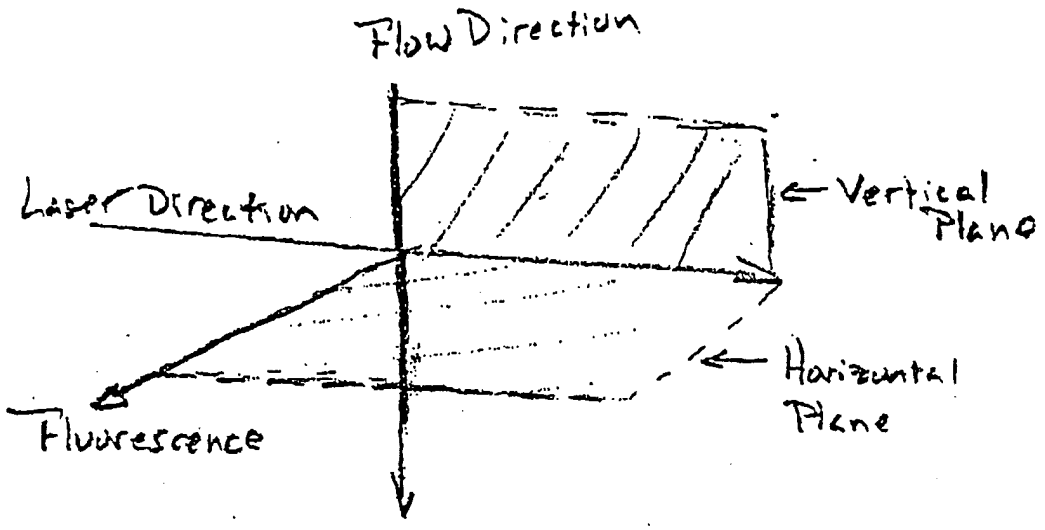

Flow Direction
Laser Direction
Vertical Plane
Horizontal Plane
Fluorescence

FIGURE 2

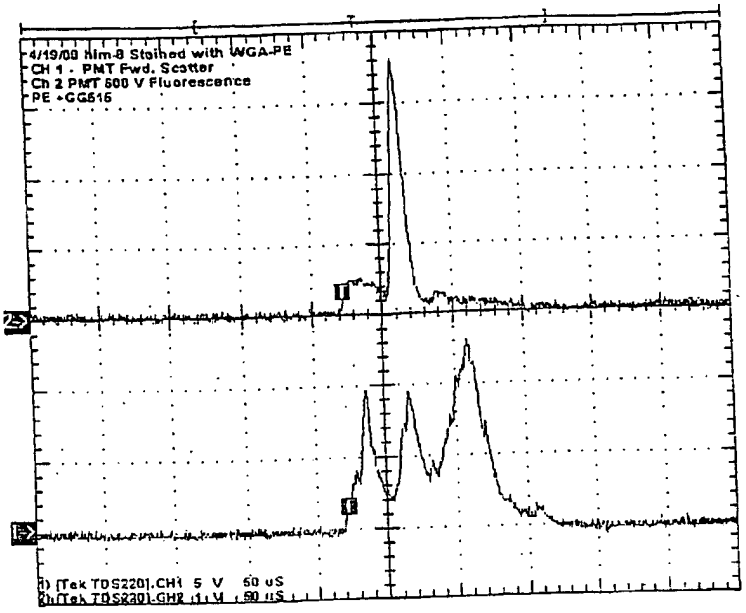


Figure 3A

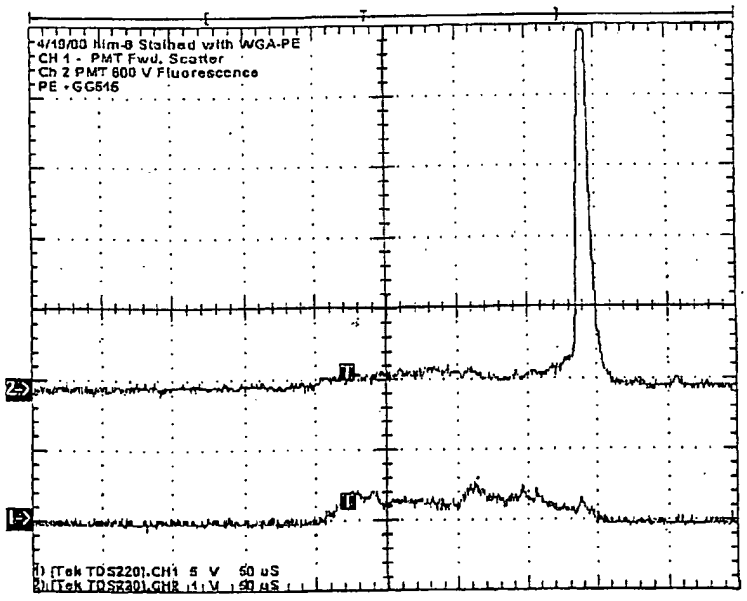


Figure 3B

Datasheet: YTSheet(1) Page: 1

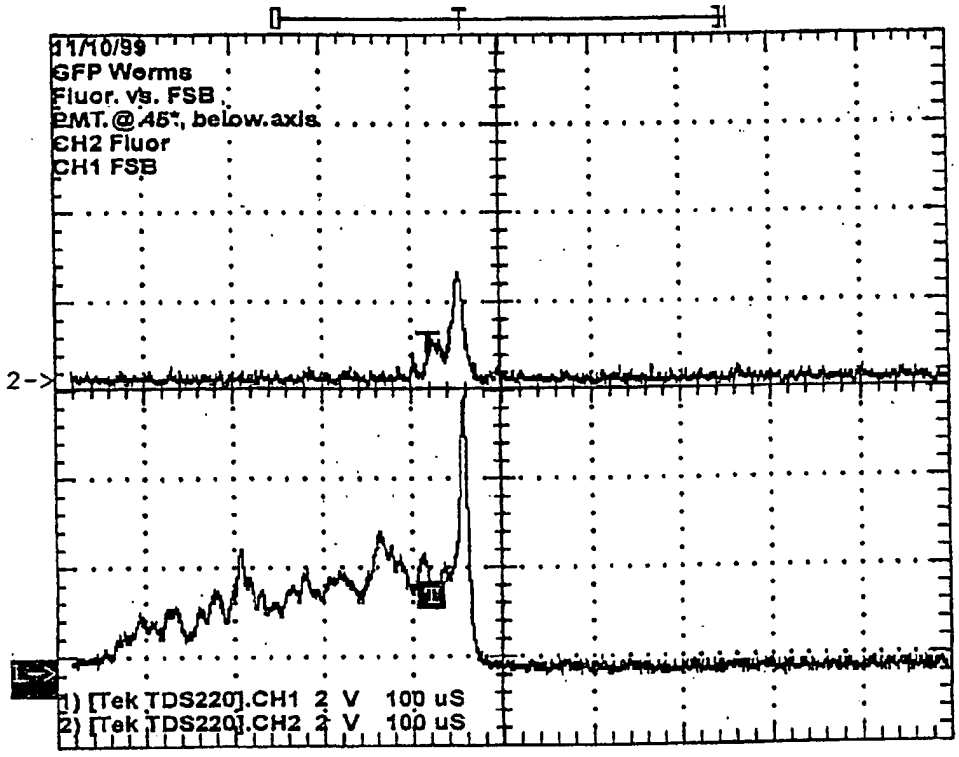


FIGURE 4

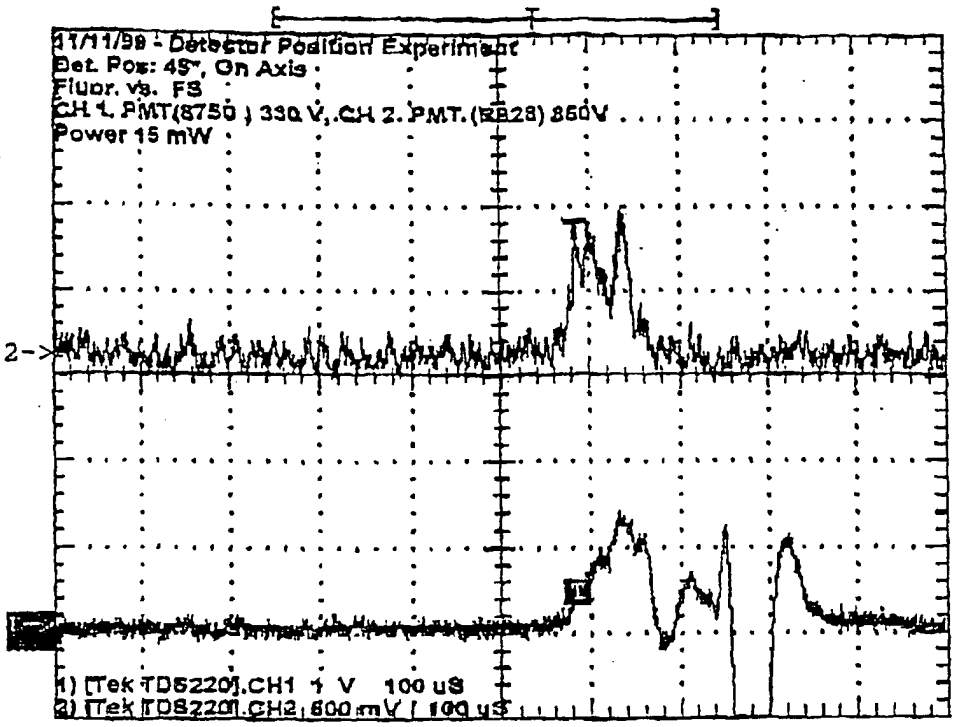

11/11/99 - Detector Position Experiment
Det. Pos: 45°, On Axis
Fluor. vs. FS
CH 1. PMT(8750) 330 V, CH 2. PMT.(R928) 850V
Power 15 mW
2->
1) Tek TDS220[.CH1 1 V 100 uS
2) Tek TDS220[.CH2 500 mV [100 uS

FIGURE 5

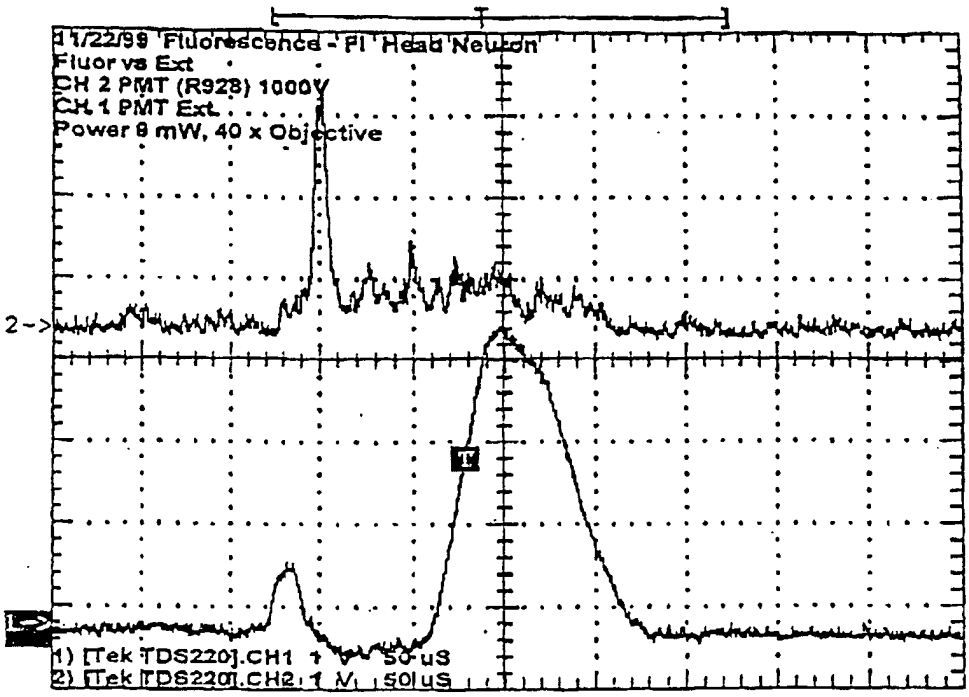

11/22/99 Fluorescence - FI Head Neuron
Fluor vs Ext
CH 2 PMT (R928) 1000V
CH 1 PMT Ext
Power 9 mW, 40 x Objective
2->
1) [Tek TDS220].CH1 1 V 50 uS
2) [Tek TDS220].CH2 1 V 50 uS

FIGURE 6

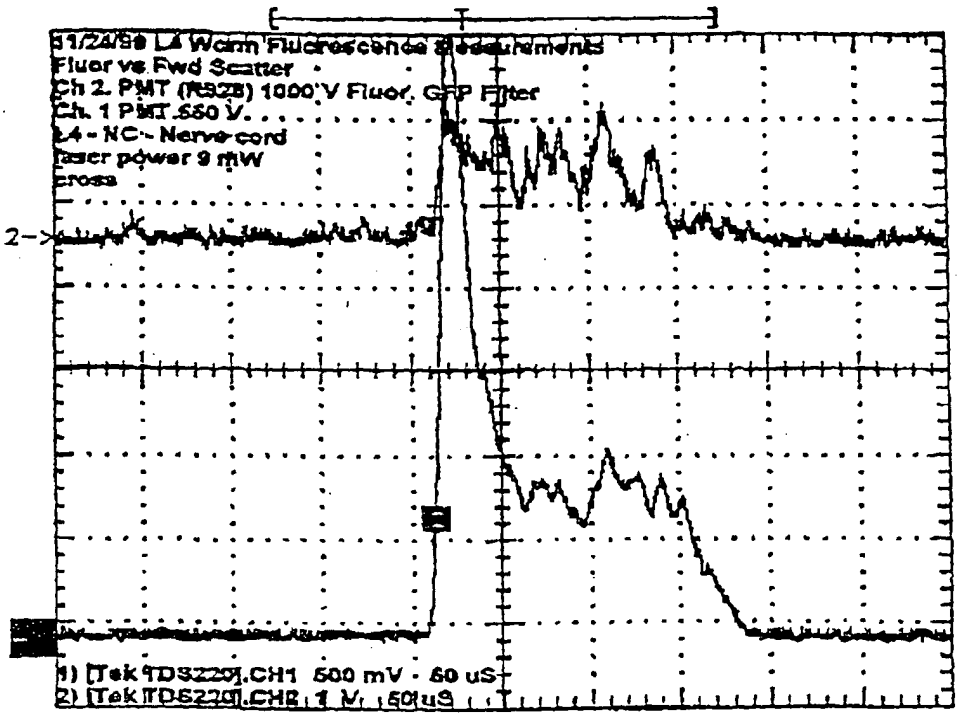

Fluor vs Fwd Scatter
Ch 2. PMT (R928) 1000 V Fluor. GFP Filter
Ch. 1 PMT 550 V.
L4 - NC - Nerve cord
laser power 9 mW
cross
2->
1) [Tek TDS220].CH1 500 mV · 50 uS
2) [Tek TDS220].CH2 1 V · 50 uS

FIGURE 7

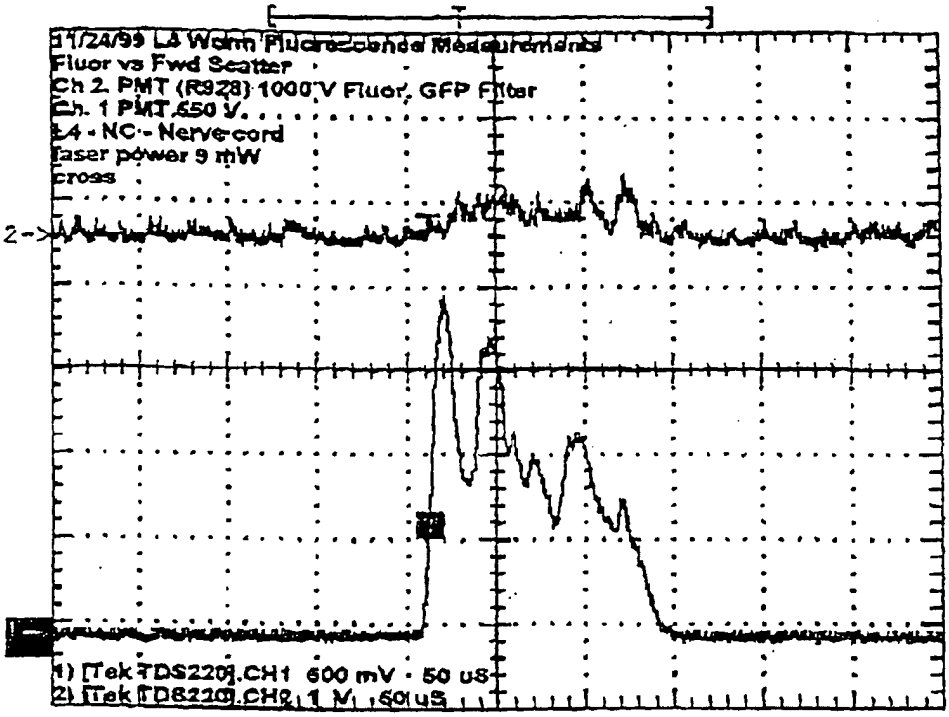

11/24/99 L3 Worm Fluorescence Measurements
Fluor vs Fwd Scatter
Ch 2. PMT (R928) 1000 V Fluor, GFP Filter
Ch. 1 PMT 650 V.
L4 - NC - Nerve cord
laser power 9 mW
cross
2->
1) [Tek TDS220].CH1 500 mV 50 uS
2) [Tek TDS220].CH2 1 V 50 uS

FIGURE 8

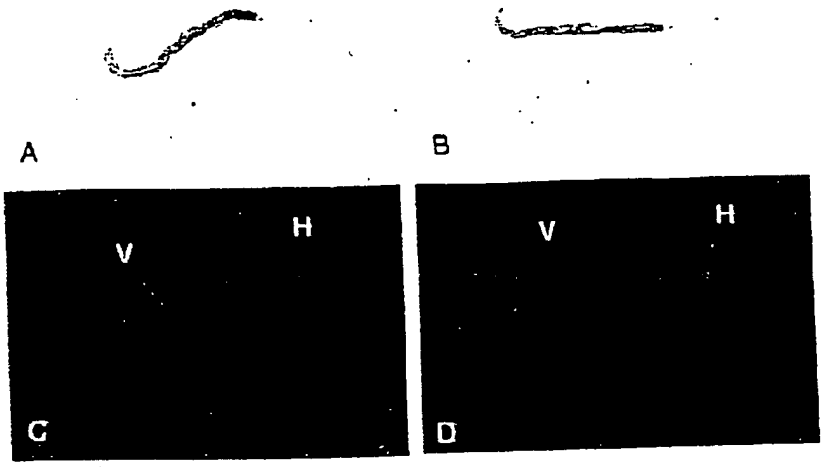


FIGURE 9

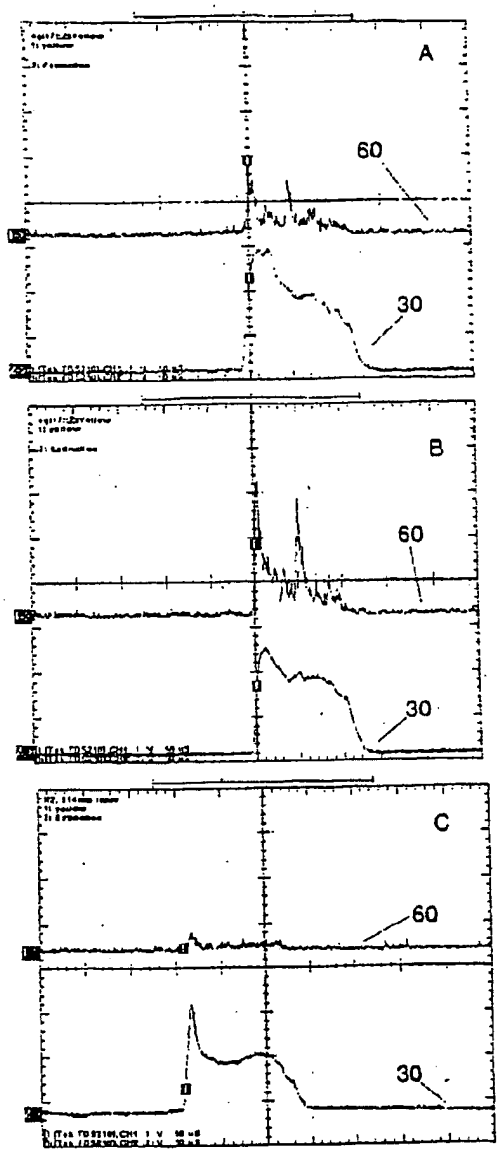


FIGURE 10

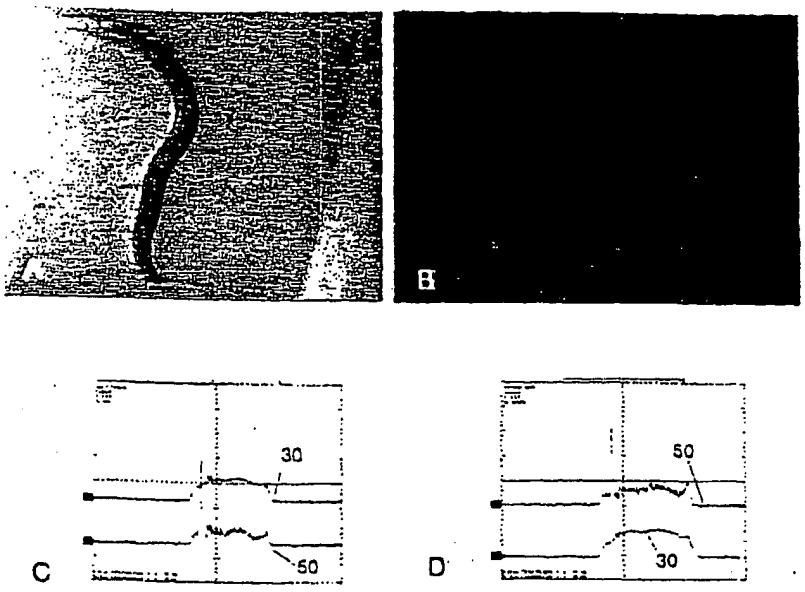


FIGURE 11

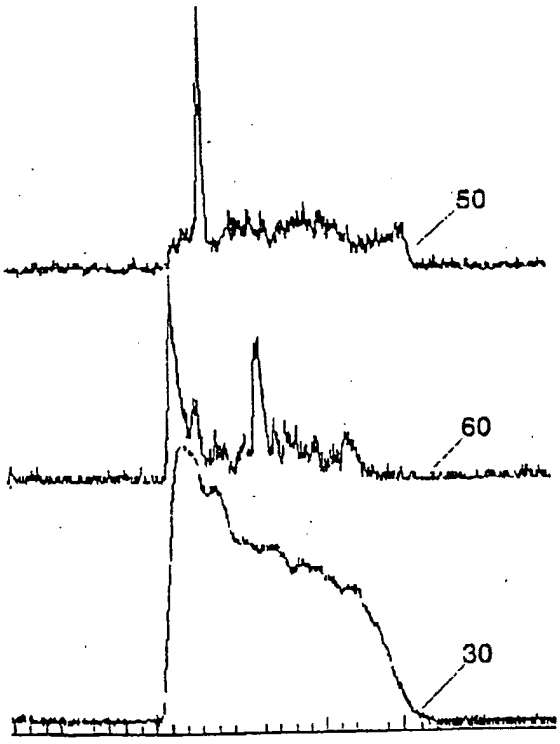


FIGURE 12

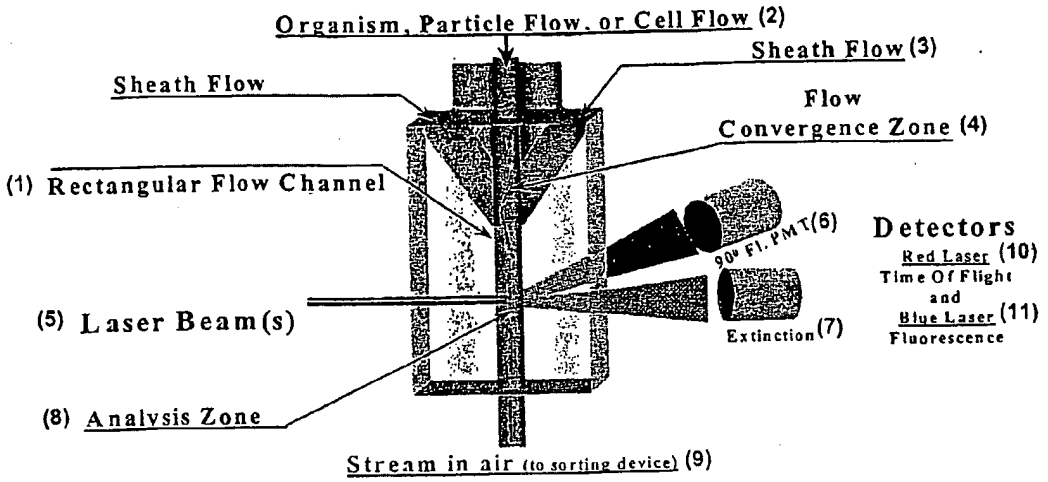

Organism, Particle Flow, or Cell Flow (2)
Sheath Flow (3)
Sheath Flow
Flow
Convergence Zone (4)
(1) Rectangular Flow Channel
90° Fl. PMT (6)
Detectors
Red Laser (10)
Time Of Flight
and
Blue Laser (11)
Fluorescence
(5) Laser Beam(s)
Extinction (7)
(8) Analysis Zone
Stream in air (to sorting device) (9)

FIGURE 13

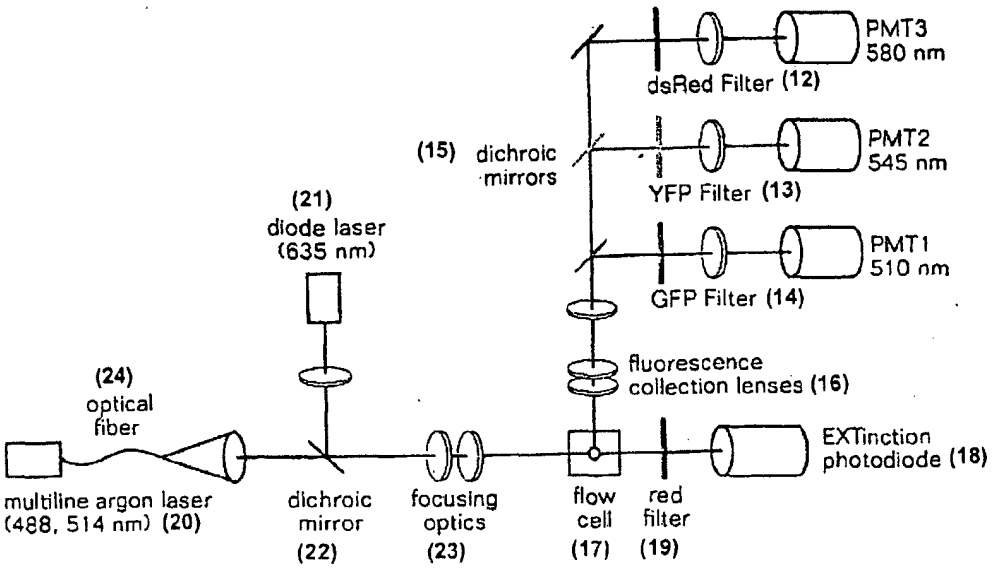

PMT3
580 nm
dsRed Filter (12)
(15) dichroic mirrors
PMT2
545 nm
YFP Filter (13)
(21)
diode laser
(635 nm)
PMT1
510 nm
GFP Filter (14)
fluorescence
collection lenses (16)
(24)
optical
fiber
EXTinction
photodiode (18)
multiline argon laser
(488, 514 nm) (20)
dichroic
mirror
(22)
focusing
optics
(23)
flow
cell
(17)
red
filter
(19)

FIGURE 14

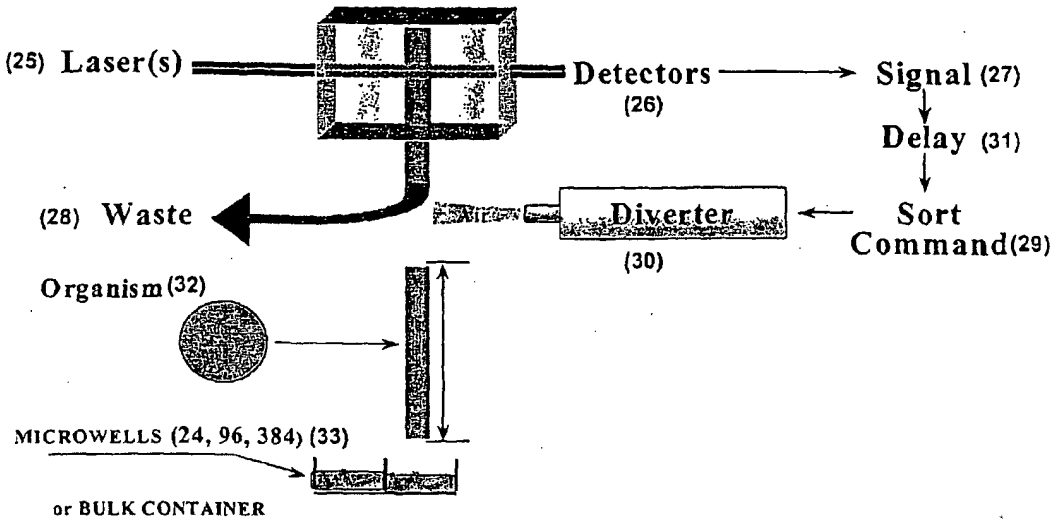

(25) Laser(s)
Detectors
(26)
Signal (27)
Delay (31)
(28) Waste
Diverter
(30)
Sort
Command(29)
Organism (32)
MICROWELLS (24, 96, 384) (33)
or BULK CONTAINER

FIGURE 15

Additional Axial Pattern Figures

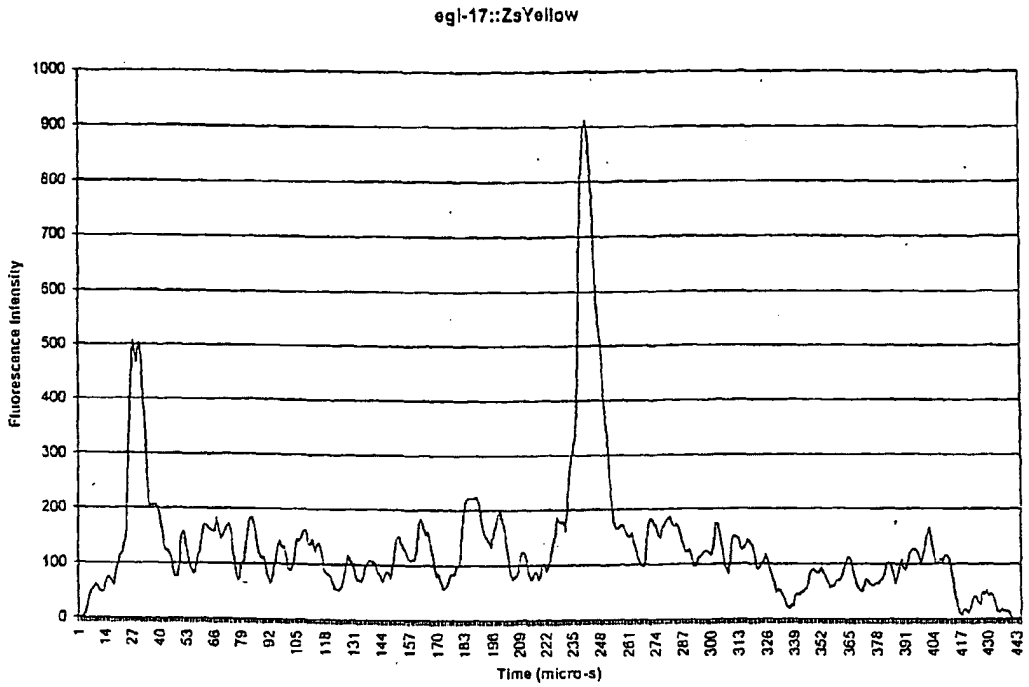


FIGURE 16

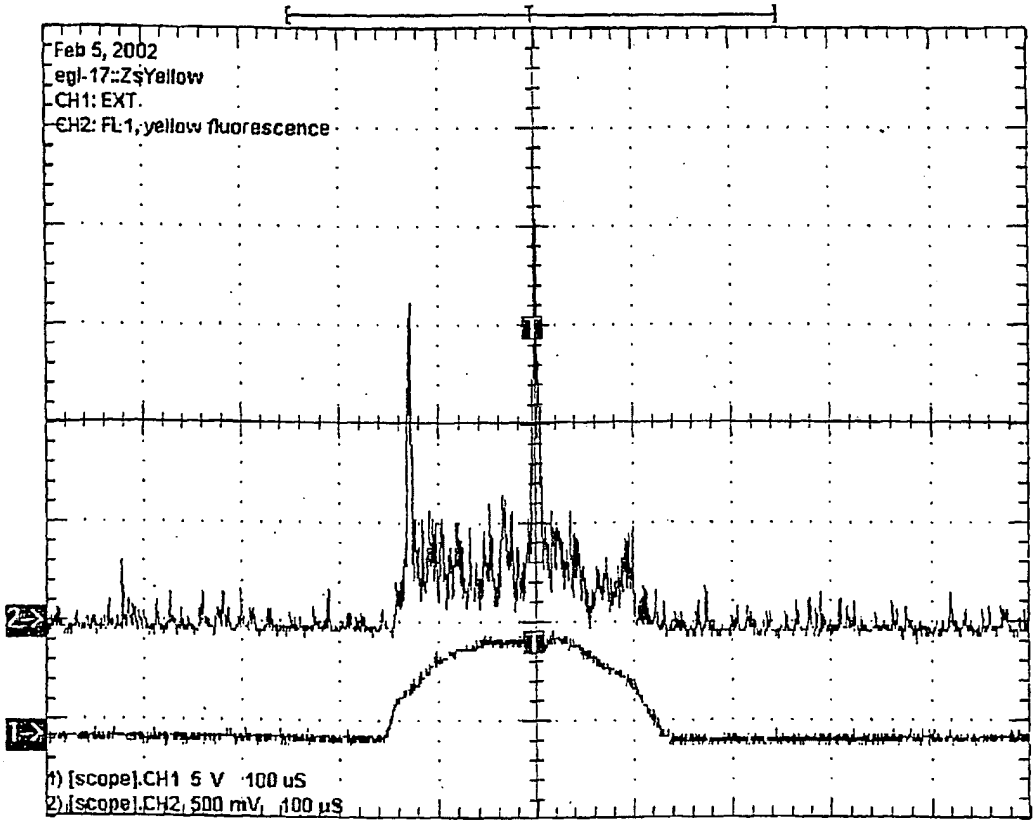

Feb 5, 2002
egl-17::ZsYellow
CH1: EXT.
CH2: FL1, yellow fluorescence
1) [scope].CH1 5 V 100 uS
2) [scope].CH2, 500 mV, 100 uS

FIGURE 17

FIGURE 18

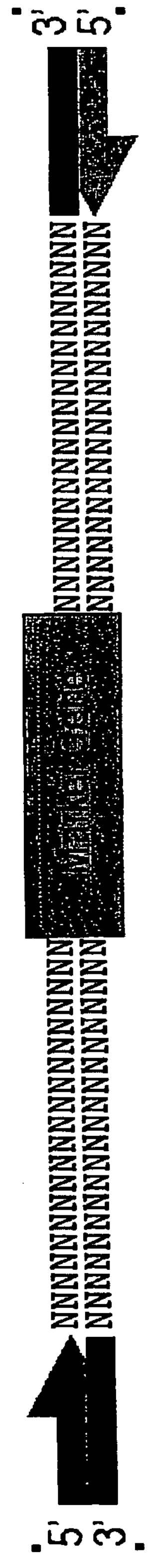


FIGURE 19

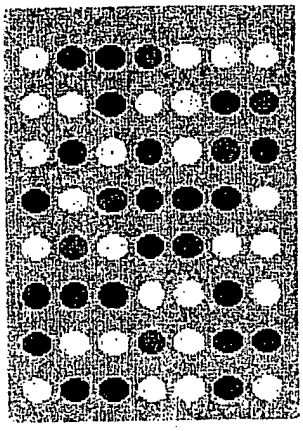


FIGURE 20